EXTENSIONS OF NONPARAMETRIC RANDOMIZATION-BASED ANALYSIS OF COVARIANCE

Michael A. Hussey

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Public Health in the Department of Biostatistics.

Chapel Hill
2013

Approved by:

Gary G. Koch
Amy H. Herring
Anastasia Ivanova
John S. Preisser, Jr.
Benjamin R. Saville
Melissa A. Troester

Abstract

MICHAEL A. HUSSEY: Extensions of Nonparametric Randomization-Based Analysis of Covariance (Under the direction of Gary G. Koch)

Nonparametric Randomization-Based Analysis of Covariance (Koch et. al. (1998)) provides covariate-adjusted estimates of treatment effects for randomized clinical trials. It has application in the regulatory setting where analyses are specified *a priori*, and any statistical assumptions of parametric methods are not verifiable until after data collection. Using (1) a vector containing differences in means of outcomes and differences in means of baseline covariables between the two randomized groups and (2) an empirical covariance matrix for the vector, weighted least squares is applied to force the difference in means of baseline covariables to zero (as expected with valid randomization) to obtain covariate-adjusted estimates. The covariate-adjusted estimates have a population-averaged interpretation and only require a valid randomization and adequate sample size (for approximate normal distributions). Saville and Koch (2012) have developed methodology combining cross-products of DFBETA residuals from a treatment-only model with covariate information to obtain a covariance matrix for use in the nonparametric covariate adjustment.

For this research, the methodology is extended to analysis of matched sets with a dichotomous outcome. In the 1:1 setting with randomization or M:1 setting, methods are provided for obtaining an adjusted difference in proportions and an adjusted odds ratio, including techniques for obtaining an exact p-value (for the difference). Application of the methods to the 1:1 observational matched case-control study is also described.

For larger strata, the methods of Saville and Koch are expanded to obtain stratified covariate-adjusted log odds ratios (in the case of dichotomous outcomes) or stratified covariate-adjusted log hazard ratios (in the case of time-to-event outcomes).

The methods of Saville and Koch are further developed for randomization to multiple treatment groups. Methodology is provided for creation of the appropriate covariance matrix to use in the nonparametric covariance adjustment, and a strategy for accommodating a time-varying treatment effect is presented.

These methods avoid modeling assumptions for the covariates (e.g. proportional hazards, functional form) while providing increased precision for the estimated treatment effects. Their use is intended for primary analysis of a randomized clinical trial, with supportive secondary parametric analyses having application to subgroup analyses or assessment of interactions with treatment.

Acknowledgments

There are numerous people who have given me advice related to the content of this dissertation and many who have given me their support throughout the process of its creation. My utmost respect and gratitude goes to my advisor, Dr. Gary Koch. Your insight and expertise helped this project happen, and I thank you for your patience while I worked to understand the bigger research picture. I also thank you for generously providing an academic home for me in the Biometric Consulting Laboratory during my graduate career at UNC. The unique experience of being a student in the BCL is one that has been transformative, and I will always look back on my time there with fond memories. Most importantly, thank you for being a mentor in the truest sense of the word. Your devotion to the next generation of biostatisticians is something I can only hope to emulate in my career.

Additionally, I want to thank the members of my committee who have generously given their time and experience to the research process: Dr. Amy Herring, Dr. Anastasia Ivanova, Dr. Robert Millikan, Dr. John Preisser, Dr. Ben Saville, and Dr. Melissa Troester. You gave me things to consider, and your contributions have improved the final product significantly.

The experience of being a doctoral student is one that is often more about endurance than it is about the difficulty of the work. It is in this regard that I have many people to thank for their support. Thank you, Mom and Dad, for always being there when I needed to talk, especially when I needed motivation to keep going. I love you guys

very much, and I am so appreciative that you have supported me in whatever career choices I've made. Many thanks to my family and friends, especially Michela Osborn, Tania Osborn, Annie Green Howard, Margaret Polinkovsky, and Ashley Pinckney. You provided balance to my academic life, and for this I will be always grateful (and sane).

Finally, I must thank the person who drew the short straw and got to be around me every day while I finished this dissertation. Michael, thank you for being a trooper through this process. Your support and motivation was essential when I wasn't sure where to go next. I love you, and I promise to never go to graduate school again.

Table of Contents

List of Tables x

List of Figures xii

1 Introduction and Literature Review 1

1.1 Development of General Nonparametric Randomization-Based ANCOVA 2

1.2 Applications to Dichotomous, Ordinal, and Survival outcomes 6

1.3 Stratification 9

1.4 Other Applications 10

1.5 Hybrid Methdology Using Diagnostics from Parametric Regression Models 12

1.6 A Note on Missing Data 13

1.7 Summary of Research 13

2 Analysis of Matched Studies with Dichotomous Outcomes using Nonparametric Randomization-Based Analysis of Covariance 15

2.1 Introduction 15

2.2 Methods 17

2.2.1 All 1:1 Matched Sets with Randomization For Paired Difference . 17

2.2.2 Informative 1:1 Matched Sets for Odds Ratio 20

2.2.3 M:1 Matched Sets with Randomization 22

2.2.4 1:1 Matched Case-Control Studies 25

2.3 Examples 26

2.3.1 Multi-Center 1:1 Clinical Trial Data for Improvement of a Skin Condition . . . 26
2.3.2 Extension of Multi-Center Clinical Trial Data to 2:1 Setting . . . 30
2.3.3 1:1 Matched Case-Control Study of Vaccine Exposure and Influenza 35
2.4 Discussion . . . 38

3 Nonparametric Randomization-Based Covariate Adjustment for Stratified Analysis of Time-to-Event or Dichotomous Outcomes . . . 40
3.1 Introduction . . . 40
3.2 Methods . . . 43
3.2.1 Hybrid Methodology for Dichotomous Outcomes with Stratification 43
3.2.2 Hybrid Methodology for Time-to-Event Outcomes with Stratification 46
3.3 Examples . . . 48
3.3.1 Dichotomous Outcome: Neurologic Disorder Data . . . 48
3.3.2 Time-to-Event Outcome: Neurologic Disorder Data . . . 50
3.3.3 Dichotomous Outcome: Osteoarthritis Data . . . 53
3.4 Discussion . . . 55

4 Covariate-adjusted Log Hazard Ratios using Cox Proportional Hazards Regression and Nonparametric Randomization-Based ANCOVA for More Than Two Treatments . . . 64
4.1 Introduction . . . 64
4.2 Methods . . . 66
4.2.1 Methodology for More than Two Treatment Groups and One Time-to-Event Outcome . . . 66
4.2.2 Estimation of Log Hazard Ratios for Time-Varying Treatment Effects with More than Two Treatment Groups . . . 70
4.3 Example . . . 72
4.3.1 Comparison of Four Treatment Groups: Neurologic Disorder Data 73

4.3.2 Comparison of Four Treatment Groups with Time-Varying Treatment Effects . . . 75

4.4 Discussion . . . 77

5 Summary and Future Research . . . 83

5.1 Summary . . . 83

5.2 Considerations and Limitations . . . 84

5.2.1 Missing Data . . . 84

5.2.2 Limitations . . . 85

5.3 Future Research . . . 85

Appendix A: Chapter 3 . . . 89

A.1 Specification of Covariance Matrix $\tilde{\boldsymbol{V}}_{\boldsymbol{f}}$ for Compound Vector in Stratified Analysis using Multivariate U-statistics . . . 89

A.1.1 Theoretical Specification . . . 89

A.1.2 Numerical Assessment . . . 93

A.2 Specification of Covariance Matrix $\bar{\boldsymbol{V}}_{\boldsymbol{f}}$ for Compound Vector in Stratified Analysis through Transformation . . . 98

A.3 Numerical Examples for Different Covariance Specifications . . . 98

A.3.1 Dichotomous Outcome: Neurologic Disorder Data . . . 99

A.3.2 Time-to-Event Outcome: Neurologic Disorder Data . . . 100

A.3.3 Dichotomous Outcome: Osteoarthritis Data . . . 101

Appendix B: Chapter 4 . . . 106

B.1 Specification of Covariance Matrix for Compound Vector in NPANCOVA for Multiple Treatment Groups . . . 106

B.2 Covariance Matrices for Multivariate Time-to-Event Outcomes . . . 110

Bibliography . . . 113

List of Tables

2.1 Matched Pairs Analysis of Multi-Center Clinical Trial Data from Example 2.3.1 . . . 28

2.2 Analysis of Multi-Center 2:1 and 1:2 Clinical Trial Data from Example 2.3.2 33

2.3 Analysis of 1:1 Matched Case-Control Study From Example 2.3.3 37

3.1 Progression by 12 Months for Patients with a Neurologic Disorder from Example 3.3.1 who Received Test Treatment (100 mg or 200 mg) or Placebo 49

3.2 Stratified Analysis for Dichotomized Outcome from Example 3.3.1 for Neurologic Disorder Data . . . 59

3.3 Strata Sizes and Number of Events for the 722 Patients Enrolled on the Placebo, 100 mg, or 200 mg Arms of the Neurologic Disorder Trial for Example 3.3.2 . . . 60

3.4 Stratified Analysis for Time-to-Event Outcome from Example 3.3.2 for Neurologic Disorder Data . . . 61

3.5 Strata Sizes for the 218 Patients with Data for Visits 1, 2, and 3 in the Osteoarthritis Randomized Crossover Trial for Example 3.3.3 . . . 62

3.6 Results from Example 3.3.3 for Osteoarthritis Data for the WOMAC Dichotomized Outcome . . . 63

4.1 Results from Analysis on Four Treatment Groups with No Time-Varying Treatment . . . 80

4.2 Parameter Estimates for Interactions from Analysis with Time-Varying Treatment . . . 81

4.3 Results from Analysis with Time-Varying Treatment . . . 82

A.3.1 Stratified Analysis for Dichotomized Outcome from Example 3.3.1 for Neurologic Disorder Data . . . 103

A.3.2 Stratified Analysis for Time-to-Event Outcome from Example 3.3.2 for Neurologic Disorder Data 104

A.3.3 Results from Example 3.3.3 for Osteoarthritis Data for the WOMAC Dichotomized Outcome 105

List of Figures

2.1 Frequencies of Success Outcomes for Treatment (T) * Placebo (P) for 1:2 and 2:1 Matched Sets for Data from Example 2.3.2 34

3.1 Kaplan-Meier Estimates of Progression-Free Survival for Neurologic Data from Example 3.3.2 51

4.1 Progression-Free Survival for Neurologic Data from Example 4.3.1 with Four Treatment Groups 73

Chapter 1

Introduction and Literature Review

Analysis of Covariance (ANCOVA) has long been used as a way of assessing the relationship between a variable and an outcome measure in the presence of other variables. In an observational study, ANCOVA may be used when there is interest in assessing the risk of disease associated with an exposure while adjusting for potential confounders. Covariance adjustment is desired in this setting so as to reduce bias in measuring the association between the exposure and risk of disease. In the randomized clinical trial setting, ANCOVA may be used to assess the effect of a treatment on an outcome measured after the treatment period while adjusting for a subject's baseline measurement. Covariance adjustment is desired in this setting so as to increase the precision of the estimated treatment effect and increase statistical power of hypothesis tests.

In general, covariance adjustment is desirable when analysis goals may involve (1) the reduction of bias due to potential confounding in an observational study, (2) the reduction of variance of an estimated treatment effect in a randomized clinical trial by adjusting for highly predictive covariables, (3) correcting minor imbalances in the distributions of baseline variables between the treatment groups which may have occurred during randomization, (4) clarification of whether statistical significance of unadjusted

treatment effects would otherwise be explained by other factors, and (5) through interaction, assessment of possible heterogeneity of treatment effects across groups defined by other factors (Snedocor and Cochran, 1980).

For continuous outcomes, parametric ANCOVA methods such as linear regression rely on the assumptions of independence of observations, homogeneity of error variances, correct functional form for covariates, and normality of the errors. For dichotomous outcomes, logistic regression requires assumptions of independent binomial distributions for the different subpopulations defined by cross-classification of the covariates as well as correct functional forms for the covariates. For ordinal and time-to-event outcomes, the proportional odds model and the Cox proportional hazards model require their respective assumptions of proportional odds and proportional hazards. In the setting of a randomized clinical trial for an experimental treatment in support of approval by a regulatory agency, all analyses on the primary endpoint must be specified prior to data collection. The statistical assumptions related to parametric ANCOVA methods are usually unverifiable *a priori*. This dilemma has led to interest in applying nonparametric randomization-based ANCOVA methods to data from randomized clinical trials.

1.1 Development of General Nonparametric Randomization-Based ANCOVA

Details for general methodology of nonparametric randomization-based ANCOVA can be found in Koch, Amara, *et. al.* (Koch et al., 1982) as well as the appendix of Koch, Tangen, *et. al.* (Koch et al., 1998). Let y_{ik} denote a univariate response for patient k randomized to treatment i in a clinical trial. Here, $k = 1, \ldots, n_i$ and $i = 1, 2$. Additionally, let $\boldsymbol{x}_{ik} = (x_{ik1}, \ldots, x_{ikP})'$ denote the vector of P covariables for the kth patient on treatment i. Then the sample mean of the responses and the sample

mean of the covariates for patients on treatment i are given as $\bar{y}_i = \frac{1}{n_i}\sum_{k=1}^{n_i} y_{ik}$ and $\bar{\boldsymbol{x}}_i = \frac{1}{n_i}\sum_{k=1}^{n_i} \boldsymbol{x}_{ik}$. Further, let $\boldsymbol{f} = (d_y, \boldsymbol{d}_x')'$ where $d_y = (\bar{y}_1 - \bar{y}_2)$ and $\boldsymbol{d}_x = (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)$. Thus $\boldsymbol{f}$ is a vector containing the difference in mean responses and the difference in means of covariables between the two treatment groups. In the case of a continuous outcome, $\boldsymbol{f}$ exists as stated. In the case of other types outcomes, the d_y may undergo transformation or represent a difference in mean scores rather than a strict difference in sample means.

A covariance matrix for $\boldsymbol{f}$ can be generated in one of two ways. Under a null hypothesis of no treatment effect, a randomization distribution can be derived under an expectation that a patient's response would be the same on either treatment. The covariance matrix that is created, $\boldsymbol{V}_0$, has the form in (1.1).

$$\boldsymbol{V_0} = \frac{n}{n_1 n_2 (n-1)} \sum_{k=1}^{n} \begin{bmatrix} (y_{*k} - \bar{y}_{**})^2 & (y_{*k} - \bar{y}_{**})(\boldsymbol{x}_k - \bar{\boldsymbol{x}}_*)' \\ (y_{*k} - \bar{y}_{**})(\boldsymbol{x}_k - \bar{\boldsymbol{x}}_*)' & (\boldsymbol{x}_k - \bar{\boldsymbol{x}}_*)(\boldsymbol{x}_k - \bar{\boldsymbol{x}}_*)' \end{bmatrix} \tag{1.1}$$

Here, $n = n_1 + n_2$, y_{*k} is the response under the null hypothesis for patient k (irrespective of treatment assignment), $\bar{y}_{**}$ is the sample mean of all patients in the trial under the null hypothesis, and $\bar{\boldsymbol{x}}_*$ is the sample mean vector of covariates for all patients in the trial. Since this covariance matrix is developed using a randomization distribution based on a finite sample (i.e. those enrolled in the clinical trial), inference is confined to the population of patients enrolled on the trial.

Alternatively, a covariance matrix can be created under the assumption that the trial patients are a simple random sample from a larger population. Then the covariance $\boldsymbol{V_s}$ can be defined as in (1.2):

$$\boldsymbol{V_s} = \sum_{i=1}^{2} \frac{1}{n_i(n_i-1)} \sum_{k=1}^{n_i} \begin{bmatrix} (y_{ik} - \bar{y}_i)^2 & (y_{ik} - \bar{y}_i)(\boldsymbol{x}_k - \bar{\boldsymbol{x}}_i)' \\ (y_{ik} - \bar{y}_i)(\boldsymbol{x}_k - \bar{\boldsymbol{x}}_i)' & (\boldsymbol{x}_k - \bar{\boldsymbol{x}}_i)(\boldsymbol{x}_k - \bar{\boldsymbol{x}}_i)' \end{bmatrix} \tag{1.2}$$

This covariance matrix allows inference to the larger population from which the trial patients were sampled, thus allowing for population-based (i.e. unconditional) inferences after covariate adjustment.

The linear model $\boldsymbol{f} = \boldsymbol{Z}b$ can be fit using weighted least squares methodology, where $\boldsymbol{Z} = (1\ \mathbf{0}_P')'$ with $\mathbf{0}_P$ representing a zero vector of length P. This model specification forces the difference in means of the covariables to zero, which is what is expected under the assumption of a valid randomization. An estimator for b can then be obtained as $b = (\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{f}$ where $\boldsymbol{V}$ is either $\boldsymbol{V_0}$ or $\boldsymbol{V_s}$. An estimator for the variance of b, $\boldsymbol{V}_b$, can be obtained as $\boldsymbol{V}_b = (\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z})^{-1}$. This estimator corresponds to the exact variance of the randomization distribution of b if $\boldsymbol{V} = \boldsymbol{V_0}$. Otherwise, the estimator corresponds to a consistent estimator of the variance of b if $\boldsymbol{V_s}$ is used.

Test statistics can be formed for the comparison of the two treatments using b and its variance estimate $\boldsymbol{V}_b$. When sample sizes are large enough for b to have an approximate normal distribution (i.e. each treatment group has at least $15\sqrt{P+1}$ patients), $Q_b = b^2/\boldsymbol{V}_b$ will have an approximate chi-square distribution with 1 degree of freedom (Koch et al., 1998). If $\boldsymbol{V_s}$ is used in the weighted least squares analysis, large-sample confidence intervals for the treatment difference can be obtained through the use of b and $\boldsymbol{V}_b$. For small samples, an exact p-value for the test of the treatment effect can be obtained (through a permutation test) when $\boldsymbol{V_0}$ is used as the covariance matrix of $\boldsymbol{f}$. Since the weighted least squares analysis using $\boldsymbol{Z}$ forces the difference in the means of the covariables between the two groups to zero, it may be of interest to assess the extent of random imbalances in the covariables from randomization. Using the $\boldsymbol{V_0}$ covariance matrix, one can create the test statistic $Q_0 = \boldsymbol{d}_x'\boldsymbol{V}_{xx}^{-1}\boldsymbol{d}_x$ which will have P degrees of freedom. Here, $\boldsymbol{V}_{xx}$ is the partition of $\boldsymbol{V}$ representing the covariance matrix of the vector of mean differences in the covariables. A non-significant p-value for the test of Q_0 would suggest goodness-of-fit in the sense that there is no evidence

of a violation of the assumption of a valid randomization.

The methodology described above is a general framework for producing nonparametric covariate-adjusted estimates of a treatment effect in a randomized clinical trial. The main assumption is that there is a valid randomization to the treatment groups so that the distributions of covariables are reasonably similar across the two treatment groups. In the analysis described above, weighted least squares forces the difference in means of the covariables between the two treatment groups to zero in line with expected differences under a valid randomization. With large enough sample sizes, the covariate-adjusted estimates will have approximate normal distributions. The choice of the covariance matrix used in the analysis will affect whether inference is restricted to the study population or whether inference pertains to a large population from which the study participants were randomly sampled.

There are some limitations to the methodology described above. Unlike standard parametric regression approaches, estimates for the association between the covariables and the response are not obtained. Usually this is not of concern, as the treatment effect is generally the main association of interest in a randomized trial. Additionally, as with any analysis of covariance, covariables should be chosen *a priori* so as to avoid selection bias from choosing adjustment variables based on their relationship to the outcome. Finally, the methodology presented here does not accommodate interactions of covariates with treatment. However, estimation of treatment effects by levels of an additional variable may be suitable in a supportive parametric analysis after statistical significance of the treatment effect is established through the nonparametric analysis.

The estimated treatment effect obtained from the nonparametric ANCOVA here should also be emphasized as representing a population-averaged treatment effect. In parametric regression approaches, the estimated treatment effect is conditional in that it only applies within subpopulations which share the same values of the covariables. Only

in the unadjusted (treatment-only) parametric model would the estimated treatment effect be population-averaged. Thus, the nonparametric ANCOVA methods tend to be better suited to situations where interest is in an overall treatment effect rather than one applying to certain subgroups of patients. It has been the recommendation of authors (Koch et al., 1998),(LaVange, Durham and Koch, 2005) that the nonparametric ANCOVA methods be reserved for analyzing the primary endpoint of a clinical trial and that parametric regression models be supportive in secondary exploratory analyses seeking to assess which subgroups may show greater benefit.

Given the flexibility of the above methodology to a variety of statistical outcomes, much research has been conducted in recent years involving the application of the nonparametric ANCOVA methods to different settings. The following sections detail extensions of the methods to dichotomous, ordinal, and survival outcomes, as well as discussing stratification and a hybrid methodology which uses parametric regression diagnostics to obtain a covariance matrix for use in a nonparametric covariance adjustment.

1.2 Applications to Dichotomous, Ordinal, and Survival outcomes

The extension of nonparametric randomization-based ANCOVA to dichotomous and ordinal outcomes is presented in Koch, Tangen, *et. al.* (Koch et al., 1998) and Tangen and Koch (Tangen and Koch, 1999*a*). One of the main benefits of covariance analysis is to achieve variance reduction in the estimated treatment effect. However, due to the nonlinearity of the logit functions in logistic or proportional odds regression, the estimated standard error of the treatment effect may increase after covariance adjustment. This may cause some concern as to whether the parametric modeling assumptions are satisfied. To avoid this concern, one can apply the nonparametric

randomization-based ANCOVA methodology to a simple transformation of the outcome measures.

Let y_{ik} be a dichotomous outcome where $y_{ik} = 1$ if patient k ($k = 1, \ldots, n_i$) shows a particular response of interest on treatment i ($i = 1, 2$) and $y_{ik} = 0$ if the patient does not show that response. Then the sample mean of the responses for each treatment is the proportion $p_i = \bar{y}_i$. The vector of covariables for adjustment follows the same structure as in the general case. One can form a vector $\boldsymbol{f}_i$ for each treatment group as $\boldsymbol{f}_i = (logit(p_i), \bar{\boldsymbol{x}}_i')'$, where $logit(p) = log_e(p/(1-p))$. A difference vector $\boldsymbol{d} = \boldsymbol{f}_1 - \boldsymbol{f}_2$ can be created which includes the difference in the log odds of response between the two treatments as well as differences in the means of the covariables between the two treatment groups. The weighted least squares methodology can then be applied to $\boldsymbol{d}$ with the model $\boldsymbol{Z} = (1 \ \boldsymbol{0}_P')'$. The covariance matrix for $\boldsymbol{d}$ used in weighted least squares is usually $\boldsymbol{V_s} = \boldsymbol{V}_1 + \boldsymbol{V}_2$ since the vector $\boldsymbol{d}$ is a vector of differences between the randomized groups. However, because a logit transformation was applied to the response portion of $\boldsymbol{d}$, the covariance matrix would have the form $\boldsymbol{D}_1\boldsymbol{V}_1\boldsymbol{D}_1 + \boldsymbol{D}_2\boldsymbol{V}_2\boldsymbol{D}_2$ where $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$ are diagonal matrices of dimension $P + 1$ where the (1,1) element of the matrices are $\frac{\partial}{\partial p_1} log_e(p_1/(1-p_1)) = \frac{1}{p_1(1-p_1)}$ and $\frac{\partial}{\partial p_2} log_e(p_2/(1-p_2)) = \frac{1}{p_2(1-p_2)}$, respectively, and 1's along the rest of the diagonal. $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$ have the same form as in (1.2) but for each treatment group separately. Chi-square tests for the treatment effect as well as random imbalances in the covariates can be conducted as described in the general methodology using the appropriate covariance structures.

In the case of the proportional odds model, one can expand the vector $\boldsymbol{d}$ to accommodate the cumulative logits. For example, if the response variable had 4 categories (e.g. 1, 2, 3, 4), then the first three entries of $\boldsymbol{d}$ would now be log odds ratios corresponding to the three cumulative logits. As usual, $\boldsymbol{d}$ would also contain the differences in the means of the covariables. Analysis could proceed as described previously with a multivariate

extension of the covariance matrix which would consider $(\boldsymbol{y}_{ik} - \bar{\boldsymbol{y}}_i)(\boldsymbol{y}_{ik} - \bar{\boldsymbol{y}}_i)'$ instead of $(y_{ik} - \bar{y}_i)^2$ and have the appropriate transformation via the relevant diagonal matrices. Assessment of the treatment effect as well as random imbalances in the covariates could proceed as previously described using $\boldsymbol{Z} = [\boldsymbol{I}_u \quad \boldsymbol{0}_{u,P}]'$ and $\boldsymbol{Z} = \boldsymbol{I}_{u+P}$, respectively. Here, u is the number of cumulative logits, $\boldsymbol{I}$ corresponds to the identity matrix, and $\boldsymbol{0}_{u,P}$ is a $u \times P$ matrix of zeroes. A test of the proportional odds assumption could proceed with the statistic $Q_w = \boldsymbol{b}'\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{V}_b\boldsymbol{C}')^{-1}\boldsymbol{C}\boldsymbol{b}$ where $\boldsymbol{C} = [-\boldsymbol{1}_{(u-1)} \quad \boldsymbol{I}_{(u-1)}]$ and $\boldsymbol{1}_{(u-1)}$ is a vector of ones of length $u-1$. Q_w is approximately chi-square with degrees of freedom corresponding to rank($\boldsymbol{C}$).

Tangen and Koch also discuss the use of nonparametric ANCOVA methodology for time-to-event data (Tangen and Koch, 1999*b*), (Tangen and Koch, 2001). One can form the vectors $\boldsymbol{f}_1$ and $\boldsymbol{f}_2$ where the first entry corresponds to the average logrank (or Wilcoxon) score for treatments 1 and 2, respectively. Bivariate testing of both types of scores could proceed with the first two entries of each vector corresponding to the averages of each type of score. The covariance matrix used for the weighted least squares portion would be based on randomization as in (1.1). Grouped survival data can be accommodated as well where time intervals are exhaustive and mutually exclusive. The response portion of the mean difference vector $\boldsymbol{d}$ consists of differences in cumulative survival rates in each of the relevant time intervals $t = 1, \ldots, T$, where the cumulative survival rates are based on products (up through the tth interval) of conditional probabilities of surviving through interval t given survival through the $(t-1)$ interval. The covariance of $\boldsymbol{d}$ is obtained through propagation of variances. A covariance matrix based on (1.2) is obtained from examining individual event and at-risk information for the time intervals jointly. A series of matrix operations then transforms this matrix to obtain the covariance matrix for $\boldsymbol{d}$. It is this matrix which can then be used in the weighted least squares covariance adjustment. P-values for the

treatment effects, a test for random imbalances, and assessment of homogeneity of the treatment effects across time intervals can all be obtained by appropriately specifying $\boldsymbol{Z}$ matrices in a manner similar to the proportional odds analysis above; however, this approach does not provide a covariate-adjusted estimate of the hazard ratio.

1.3 Stratification

The nonparametric ANCOVA methods described thus far account for factors as covariables. Extensions of the methodology to stratified studies have been presented (Koch et al., 1998),(Tangen and Koch, 1999a),(Tangen and Koch, 1999b),(LaVange, Durham and Koch, 2005). For randomized trials, a common stratification factor is center in the context of a multi-center clinical trial. Other common stratification factors could be geographic region or prognostic factors. For the purposes of nonparametric randomization-based ANCOVA, two different approaches are recommended depending on whether the strata are small (i.e. treatment group sizes within each stratum are generally ≤ 15) or large. In the case of small strata, the difference vector $\boldsymbol{d}_h$ is created for each of the q strata ($h = 1, \ldots, q$). Covariance matrices $\boldsymbol{V_{s,h}}$ are also obtained for each stratum. A weighted difference vector $\boldsymbol{d}_w = \sum_{h=1}^{q} w_h \boldsymbol{d}_h$ is obtained as well as a weighted covariance matrix $\boldsymbol{V_{s,w}} = \sum_{h=1}^{q} w_h \boldsymbol{V_{s,h}}$. Here, the w_h are usually the Mantel-Haenszel weights (Mantel and Haenszel, 1959). Weighted least squares is then applied using $\boldsymbol{d}_w$ and $\boldsymbol{V_{s,w}}$. When strata are large enough, weighted least squares can be applied within each stratum and then the estimates b_h and $V_{b,h}$ can be combined over strata using the Mantel-Haenszel weights. Associated tests for the treatment effect can be obtained as well (Koch et al., 1998).

1.4 Other Applications

Several other applications of nonparametric randomization-based ANCOVA have been developed in recent years. Koch and Tangen extend the methodology to non-inferiority clinical trials (Koch and Tangen, 1999). Consider a randomized trial with test treatment (T), reference treatment (R), and placebo (P). Interest is in making sure that both T and R are better than placebo and that T is also non-inferior to R. The difference vector $\boldsymbol{d}$ can take the form of $(d_{TP}, \boldsymbol{u}'_{TP}, d_{RP}, \boldsymbol{u}'_{RP})'$ where d_{TP} and d_{RP} are the difference in mean responses between test and placebo as well as reference and placebo, respectively. The $\boldsymbol{u}_{TP}$ and $\boldsymbol{u}_{RP}$ represent the differences in the means of the covariables for test vs. placebo as well as reference vs. placebo. The differences in the means of the covariables can be forced to zero through a weighted least squares analysis which produces covariate-adjusted estimates of the differences between each response and placebo jointly. Through an appropriate specification of $\boldsymbol{Z}$, the ratio of (T-P) relative to (R-P) can be assessed, and Fieller's formula (Fieller, 1954) can be used to obtain a confidence interval for this ratio.

Kawaguchi, Koch, and Wang extend the methodology to the case where stratified multivariate Mann-Whitney estimators are used to compare two treatments with respect to any number of finite ordinal responses (Kawaguchi, Koch and Wang, 2011). Here, the elements of vector $\boldsymbol{f}$ include the stratification-adjusted Mann-Whitney estimators for each of the responses as well as ratios comparing the two treatment groups with respect to stratification-adjusted estimators of covariable means. A consistent covariance matrix for $\boldsymbol{f}$ can be obtained through propagation of variance and then used in the usual weighted least squares analysis to obtain covariate-adjusted estimates of the treatment effect.

Saville, LaVange, and Koch extend the methodology to the estimation of incidence density ratios in multiple time intervals (Saville, Lavange and Koch, 2011). They

use example data from a clinical trial for COPD where pulmonary exacerbations were recorded as the events of interest. Considering six 6-month time intervals (3 years of potential follow-up), the elements of the vector $\boldsymbol{f}_{hik}$ are the counts of exacerbations, the time at risk, and the covariable values for individual k on treatment i in the hth stratum during the J time intervals. A subsequent covariance matrix $\boldsymbol{V_{hi}}$ can be formed for $\bar{\boldsymbol{f}}_{hi} = \sum_{k=1}^{n_i}(\boldsymbol{f}_{hik}/n_i)$. Weighted estimates of the $\bar{\boldsymbol{f}}_{hi}$ and the $\boldsymbol{V_{hi}}$ can be formed. From these, a compound difference vector $\boldsymbol{d}$ can be formed which contains the stratification-adjusted, log incidence density ratios comparing the two treatments in each of the J time intervals as well as the stratification-adjusted differences in the means of the P covariables. Weighted least squares can then be applied to $\boldsymbol{d}$ to obtain covariate-adjusted estimates of the log incidence density ratios. Subsequent hypothesis tests and confidence intervals can be formed, including the test to check for random imbalances in the covariates between the treatment groups. While the procedure just described applies an order of stratification, estimation of log incidence density ratios, and covariate adjustment, the ordering of these three steps can be changed depending on the number of time intervals or the size of the strata. If the number of time intervals is small, the covariate adjustment can occur prior to estimation of the log incidence density ratios. In the event of large strata, estimation of log incidence density ratios and covariate adjustment can occur within strata before the covariate-adjusted estimates are combined across the strata using some form of weighting.

A similar approach was used to estimate covariate-adjusted log hazard ratios for multiple time intervals as in Moodie, Saville, Koch, and Tangen (Moodie et al., 2011). Elements of the vector $\boldsymbol{f}_{ik}$ for the kth patient on the ith treatment consist of J terms which assess survival of the time intervals and J terms which assess whether the patient was at risk in the J pre-specified time intervals under study. This vector also includes the covariate values for the kth patient on the ith treatment. The mean for

each treatment $\boldsymbol{f}_i$ can be formed, and an associated covariance matrix can be constructed. Through the propagation of variance, a vector $\boldsymbol{d}$ is formed which represents the difference between the treatment groups with respect to log hazard ratios and the means of the covariables. Weighted least squares then uses $\boldsymbol{d}$ and its covariance matrix $\boldsymbol{V_d}$ to obtain covariate-adjusted estimates of the log hazard ratios for the J time intervals. As in the general case, hypothesis tests for homogeneity of the log hazard ratios from the J time intervals can be conducted along with a test of random imbalance in the covariates between the two treatment groups.

1.5 Hybrid Methdology Using Diagnostics from Parametric Regression Models

A common semiparametric approach to estimating a hazard ratio in the presence of covariables is the Cox proportional hazards model for time-to-event data (Cox, 1972). Saville and Koch proposed a covariate-adjusted log hazard ratio obtained via nonparametric randomization-based ANCOVA which is closer to that of the Cox model than what was described in Moodie *et. al.* (Moodie et al., 2011) (Saville and Koch, 2013). Marginal Cox proportional hazards models containing a treatment indicator are fit for each of K events. The K DFBETAs from these models are obtained for individuals in each of the two treatment groups. The DFBETA for patient j ($j = 1, \ldots, n_i$) and event k ($k = 1, \ldots, K$) represents the approximate change to the kth estimated log hazard ratio when the jth patient is omitted. Wei *et. al.* showed that an approximate covariance matrix for the estimated log hazard ratios $\hat{\boldsymbol{\beta}}$, $\boldsymbol{V}_{\hat{\boldsymbol{\beta}}}$, can be obtained by taking the sum of the cross-products of the DFBETAs (Wei, Lin and Weissfeld, 1989). If the elements of the difference vector $\boldsymbol{d}$ are the estimated log hazard ratios $\hat{\boldsymbol{\beta}}$ and the differences in means of the covariables $(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)$, the associated covariance matrix $\boldsymbol{V_d}$ can be assembled using $\boldsymbol{V}_{\hat{\beta}}$, $\boldsymbol{V}_{\bar{\boldsymbol{x}}_1} + \boldsymbol{V}_{\bar{\boldsymbol{x}}_2}$, and $\boldsymbol{V}_{\hat{\beta},\bar{\boldsymbol{x}}_1} - \boldsymbol{V}_{\hat{\beta},\bar{\boldsymbol{x}}_2}$. The matrix $\boldsymbol{V}_{\bar{\boldsymbol{x}}_1} + \boldsymbol{V}_{\bar{\boldsymbol{x}}_2}$

is the estimated covariance matrix for $(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)$ and $\boldsymbol{V}_{\hat{\beta},\bar{\boldsymbol{x}}_1} - \boldsymbol{V}_{\hat{\beta},\bar{\boldsymbol{x}}_2}$ is the estimated covariance of $\hat{\boldsymbol{\beta}}$ with $(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)$. Here, $\boldsymbol{V}_{\hat{\beta},\bar{\boldsymbol{x}}_i}$ is the estimated covariance matrix for $\hat{\beta}$ and $\bar{\boldsymbol{x}}_i$ for the ith treatment group. Weighted least squares can then be applied using $\boldsymbol{d}$ and $\boldsymbol{V_d}$ via the model $\boldsymbol{Z} = [\boldsymbol{I}_K \quad \boldsymbol{0}_{KP}]'$ to obtain covariate-adjusted log hazard ratios. Tests of homogeneity across the K log hazard ratios and random imbalance of covariables between the two treatment groups can be conducted per other specifications of $\boldsymbol{Z}$. While this hybrid procedure is not thought to be fully nonparametric since it makes use of a treatment-only Cox model, the p-value from the Cox model is comparable to the nonparametric logrank test (Saville and Koch, 2013). The assumption of proportional hazards for the covariates is avoided since covariate adjustment is made via the nonparametric methodology.

1.6 A Note on Missing Data

The general nonparametric randomization-based ANCOVA methodology relies on the assumption that the distributions of the covariables between the two treatment groups are comparable as a consequence of a valid randomization. Exclusions of patients for protocol violations or missingness on a covariate could potentially induce imbalance in the covariate distribution between the two groups (Koch et al., 1998). However, the methods would be applicable for any sort of valid multiple imputation strategy. For the methodology in Kawaguchi, Koch, and Wang, missing data could be accommodated when using the Mann-Whitney estimators, although a mechanism of missing completely at random (MCAR) is assumed (Kawaguchi, Koch and Wang, 2011).

1.7 Summary of Research

The following chapters outline applications of nonparametric ANCOVA methodology to novel settings. Chapter 2 is a direct application of the methodology to matched

sets data with dichotomous outcomes. Nonparametric methodology is introduced for obtaining covariate-adjusted estimates of a treatment effect for 1:1 and M:1 randomized studies as well as for 1:1 non-randomized matched case-control studies. Such applications are intended for matched pairs data (as an alternative to the unadjusted McNemar's analysis or conditional logistic regression) or when the size of the matched set is $\leq$6. Chapter 3 outlines a hybrid methodology (like that in Section 1.5) for conditional likelihoods (as in conditional logistic regression) or partial likelihoods (as in proportional hazards regression) in the presence of stratification. A treatment-only regression model is fit, and the sum of the squared DFBETA residuals estimates the covariance of the treatment effect which can then be used in a weighted least squares analysis to provide covariate adjustment. Chapter 4 extends the hybrid methodology to more than two treatment groups for time-to-event data as explored in Saville and Koch (2012) with additional methodology for a time-varying treatment effect.

Chapter 2

Analysis of Matched Studies with Dichotomous Outcomes using Nonparametric Randomization-Based Analysis of Covariance

2.1 Introduction

Matching may arise in clinical trials where the objective is to establish superiority of a new treatment to a control (which may or may not be placebo). These studies may range from a multi-center clinical trial (where subjects in the same randomization block or from the same site form a matched set) to studies in dermatology, opthalmology, or dentistry (where an individual's arms/legs, eyes, or sections of the mouth form a matched set). Matching allows for eliminating variability in the outcome among the matched sets, thus making it a useful technique when designing a clinical study. In the case of matched pairs for a dichotomous outcome, McNemar's test and its related odds ratio can be used to assess whether the subjects are more likely to experience success (vs. failure) on a new treatment than they are on the control (McNemar, 1947). If covariate adjustment is desired for baseline variables that vary within pairs,

analysis of a binary outcome where data are from matched sets may proceed using conditional logistic regression (Breslow and Day, 1980). In either type of analysis, log odds ratios and their confidence intervals can be obtained to assess the effect of the new treatment. Matched studies may also be of the form M:1, where M individuals are randomized to treatment and one individual is randomized to control within a matched set. In studies with blocked randomization, blocks formed as successive pairs, triples, or groups of $M + 1$ patients may be matched on time of enrollment if it was thought to be a prognostic factor. Retrospective case-control studies may also exhibit M:1 matching, where M controls are matched to one case on variables which may be thought to confound the relationship between an exposure variable and the odds of being a case. For most practical purposes, M ranges between two and four subjects (Stokes, Davis and Koch, 2012).

The NPANCOVA methodology is useful for providing covariate-adjusted estimates of treatment effects for data from matched studies when assumptions of other methods are not verifiable. In studies with randomization, a valid randomization is assumed so that the differences in covariate means between the randomized groups are expected to be zero. In the case of the observational matched case-control study (without randomization), the NPANCOVA methodology may be applicable provided that the exposed and not exposed subjects within the informative matched sets have similar distributions of other covariables across the matched sets if the matching is successful. This chapter presents extensions of the NPANCOVA methodology to randomized studies with 1:1 or M:1 matching as well as an observational case-control study with 1:1 matching. Examples presented include a randomized clinical trial with 1:1 matching, a randomized clinical trial with 2:1 matching, and a retrospective case-control study with 1:1 matching. Discussion of the advantages and limitations of the methodology follows the examples.

2.2 Methods

2.2.1 All 1:1 Matched Sets with Randomization For Paired Difference

Let $y_{hi} = 1$ denote a success on the ith treatment ($i = 1, 2$) for pair h ($h = 1, \ldots, q$), and let $y_{hi} = 0$ denote a failure. Additionally, let $\boldsymbol{x}_{hi}$ be the $P \times 1$ vector of P covariates corresponding to the ith treatment for pair h. Within each of the respective pairs, the two treatments have independent random allocation with equal probabilities of 0.5. A vector of the differences in binary outcomes for the two treatments can be formed as $d_h = (y_{h1} - y_{h2})$. The possible values for d_h are as follows:

$$d_h = y_{h1} - y_{h2} = \begin{cases} 1 & \text{if } y_{h1} = 1,\ y_{h2} = 0 \\ 0 & \text{if } y_{h1} = y_{h2} \\ -1 & \text{if } y_{h1} = 0, y_{h2} = 1 \end{cases}$$

With $\boldsymbol{u}_h = (\boldsymbol{x}_{h1} - \boldsymbol{x}_{h2})$ denoting the difference between treatments for the covariables $\boldsymbol{x}_{hi}$ within a pair, a vector $\boldsymbol{f}_h = (d_h, \boldsymbol{u}_h')'$ can be formed for each pair h. The respective $\boldsymbol{f}_h$ are statistically independent on the basis of (1) the respective pairs being comparable to a random sample from a relevant target population and (2) the independent random allocation of the two treatments within corresponding pairs. The mean vector $\bar{\boldsymbol{f}} = (\bar{d}, \bar{\boldsymbol{u}}')' = \frac{1}{q}\sum_{h=1}^{q} \boldsymbol{f}_h$ can be formed with an unbiased estimate for its covariance matrix $\boldsymbol{V}_{\bar{\boldsymbol{f}}}$ as shown in (2.1). Here, $\bar{d}$ represents the unadjusted difference in proportions of success between the treatments.

$$\boldsymbol{V}_{\bar{\boldsymbol{f}}} = \frac{1}{q(q-1)} \sum_{h=1}^{q} (\boldsymbol{f}_h - \bar{\boldsymbol{f}})(\boldsymbol{f}_h - \bar{\boldsymbol{f}})' = \begin{bmatrix} v_{\bar{d}} & \boldsymbol{V}'_{\bar{d},\bar{\boldsymbol{u}}} \\ \boldsymbol{V}_{\bar{d},\bar{\boldsymbol{u}}} & \boldsymbol{V}_{\bar{\boldsymbol{u}}} \end{bmatrix} \tag{2.1}$$

A covariate-adjusted estimate of the difference in proportions of successes on the treatments can then be obtained using weighted least squares methodology by forcing the difference in means for the covariables to zero. With adjustment for P covariables, the model $\boldsymbol{Z} = (1\ \mathbf{0}'_P)'$ can be fit to $\bar{\boldsymbol{f}}$, where $\mathbf{0}_P$ is a zero vector of length P. The covariate-adjusted estimate b for the difference in proportions is given in (2.2).

$$\begin{aligned} b &= (\boldsymbol{Z}'\boldsymbol{V}_{\bar{\boldsymbol{f}}}^{-1}\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{V}_{\bar{\boldsymbol{f}}}^{-1}\bar{\boldsymbol{f}} \\ &= (\bar{d} - \boldsymbol{V}'_{\bar{d},\bar{\boldsymbol{u}}}\boldsymbol{V}_{\bar{\boldsymbol{u}}}^{-1}\bar{\boldsymbol{u}}) \\ &= \sum_{h=1}^{q}(d_h - \boldsymbol{V}'_{\bar{d},\bar{\boldsymbol{u}}}\boldsymbol{V}_{\bar{\boldsymbol{u}}}^{-1}\boldsymbol{u}_h)/q \end{aligned} \tag{2.2}$$

A consistent estimator for the variance of b is V_b in (2.3).

$$\begin{aligned} V_b &= (\boldsymbol{Z}'\boldsymbol{V}_{\bar{\boldsymbol{f}}}^{-1}\boldsymbol{Z})^{-1} \\ &= (v_{\bar{d}} - \boldsymbol{V}'_{\bar{d},\bar{\boldsymbol{u}}}\boldsymbol{V}_{\bar{\boldsymbol{u}}}^{-1}\boldsymbol{V}_{\bar{d},\bar{\boldsymbol{u}}}) \\ &= \sum_{h=1}^{q}\left((d_h - \bar{d}) - \boldsymbol{V}'_{\bar{d},\bar{\boldsymbol{u}}}\boldsymbol{V}_{\bar{\boldsymbol{u}}}^{-1}(\boldsymbol{u}_h - \bar{\boldsymbol{u}})\right)^2/q(q-1) \end{aligned} \tag{2.3}$$

where $v_{\bar{d}}$, $\boldsymbol{V}_{\bar{d},\bar{\boldsymbol{u}}}$, and $\boldsymbol{V}_{\bar{\boldsymbol{u}}}$ are the sub-matrices of $\boldsymbol{V}_{\bar{\boldsymbol{f}}}$ that correspond to the variance of $\bar{d}$, the covariance vector of $\bar{d}$ with $\bar{\boldsymbol{u}}$, and the covariance matrix of $\bar{\boldsymbol{u}}$, respectively. The estimator b has an approximate normal distribution when the number of matched pairs q is sufficiently large such that $\bar{\boldsymbol{f}}$ has an approximately multivariate normal distribution. Where n_1 and n_{-1} represent the number of pairs with $d_h = 1$ and the number of pairs with $d_h = -1$, $min(n_1, n_{-1})/(P+1)$ should be ≥ 8 and, ideally, ≥ 10. It follows that the statistic b^2/V_b has an asymptotic chi-squared distribution with one degree of freedom. A $100^*(1-\alpha)\%$ confidence interval for b may be obtained as $b \pm z_{1-\alpha/2} * \sqrt{V_b}$ where $z_{1-\alpha/2}$ is the $100^*(1-\alpha/2)$th percentile of a standard normal distribution. When q is only moderately large (e.g., $25 \leq q \leq 60$), the t-distribution with $(q - q_0 - 1 - P)$

degrees of freedom (where q_0 is the number of pairs with $d_h = 0$) would improve the approximate basis of this confidence interval.

Under the null hypothesis of no treatment difference whereby each individual in a pair has the same response regardless of treatment assignment, the exact distribution for $\bar{\boldsymbol{f}}$ has $2^{(q-q_0)}$ possible realizations, and the corresponding exact covariance matrix $\boldsymbol{V}_{\bar{\boldsymbol{f}},\mathbf{0}}$ has the form in (2.4).

$$\begin{aligned} \boldsymbol{V}_{\bar{\boldsymbol{f}},\mathbf{0}} &= \frac{1}{q^2}\sum_{h=1}^{q} \boldsymbol{f}_h \boldsymbol{f}_h' \\ &= \begin{bmatrix} v_{\bar{\boldsymbol{d}},\mathbf{0}} & \boldsymbol{V'}_{\bar{d},\bar{\boldsymbol{u}},\mathbf{0}} \\ \boldsymbol{V}_{\bar{d},\bar{\boldsymbol{u}},\mathbf{0}} & \boldsymbol{V}_{\bar{\boldsymbol{u}},\mathbf{0}} \end{bmatrix} \end{aligned} \tag{2.4}$$

One then obtains the estimate $b_0 = (\boldsymbol{Z}'\boldsymbol{V}_{\bar{\boldsymbol{f}},\mathbf{0}}^{-1}\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{V}_{\bar{\boldsymbol{f}},\mathbf{0}}^{-1}\bar{\boldsymbol{f}}$, an estimator of its variance $V_{b,0} = (\boldsymbol{Z}'{\boldsymbol{V}_{\bar{\boldsymbol{f}},\mathbf{0}}}^{-1}\boldsymbol{Z})^{-1}$, and $Q_{b,0} = b_0^2/V_{b,0}$ for each of N re-randomizations of the treatment assignment. The proportion of the N re-randomizations with $Q_{b,0}$ greater than or equal to the observed $Q_{b,0}$ would be the essentially exact p-value. Due to computational feasibility, choosing N to be sufficiently large for appropriate precision (e.g., $N = 100,000$) may be a necessary Monte Carlo alternative to computing $Q_{b,0}$ for all 2^q possible treatment assignments so that the standard error of p is about 0.0005 for one-sided p-values that are near or below 0.025.

Since $\boldsymbol{V}_{\bar{\boldsymbol{f}},\mathbf{0}}$ is constant for all 2^q possible treatment assignments, one could compare each b_0 from the N re-randomizations to the observed b_0 instead of using $Q_{b,0}$ to obtain the exact p-value; the result will be identical. The exact analysis can be conducted using a procedure such as PROC MULTTEST in SAS software (SAS Institute, Cary, NC). However, it is important to note that when using the variance $\boldsymbol{V}_{\bar{\boldsymbol{f}},\mathbf{0}}$ in (2.4), a trade-off is that the formal inference is restricted to the randomized study population, and so broader generalization is informal. This posture is in contrast to the variance

$\boldsymbol{V}_{\bar{\boldsymbol{f}}}$ in (2.1) which assumes generalizability through pairs being like a random sample of pairs from a corresponding population.

This methodology relies upon the assumption of a valid randomization, under which no differences between the means of the covariables are expected between the two treatment groups. Random imbalances in these means between the two groups can be assessed with the statistic

$$Q_{\bar{\boldsymbol{u}}} = (\bar{\boldsymbol{f}} - \boldsymbol{Z}b)' \boldsymbol{V}_{\bar{\boldsymbol{f}}}^{-1} (\bar{\boldsymbol{f}} - \boldsymbol{Z}b) \tag{2.5}$$

For sufficiently large q, this statistic approximately has the chi-squared distribution with P degrees of freedom, and its counterpart $Q_{\bar{\boldsymbol{u}},0}$ with b_0 and $\boldsymbol{V}_{\bar{\boldsymbol{f}},\boldsymbol{0}}$ replacing b and $\boldsymbol{V}_{\bar{\boldsymbol{f}}}$ can have exact assessment.

2.2.2 Informative 1:1 Matched Sets for Odds Ratio

If the assumption of similar covariate distributions in the treatment groups remains applicable when non-informative pairs with $d_h = 0$ are removed from the analysis, other analysis approaches may be justifiable. Consider $a_{h,I}$ for informative pairs, where $a_{h,I} = 1$ if Treatment 1 is a success and Treatment 2 is a failure and $a_{h,I} = 0$ if Treatment 1 is a failure and Treatment 2 is a success. Also, $a_{h,I}$ is considered missing when the same outcome is observed for Treatment 1 and Treatment 2 within a pair, and such non-informative pairs h with missing $a_{h,I}$ are omitted from the analysis. With $\boldsymbol{z}_{h,I} = (a_{h,I}, \boldsymbol{u}'_{h,I})'$, $\bar{\boldsymbol{z}}_I = (\bar{a}_I, \bar{\boldsymbol{u}}'_I)' = \frac{1}{q'} \sum_{h=1}^{q'} \boldsymbol{z}_{h,I}$, and $\boldsymbol{V}_{\bar{\boldsymbol{z}}_I}$ formed in a similar manner as $\boldsymbol{V}_{\bar{\boldsymbol{f}}}$ in (2.1), the model $\boldsymbol{Z} = (1\ \boldsymbol{0}'_P)'$ can be fit to $\bar{\boldsymbol{z}}_{\boldsymbol{I}}$. Here, $\bar{a}_I$ represents the proportion of pairs (among $q' = (q - q_0)$ informative pairs) with success on Treatment 1 and failure on Treatment 2. An estimate $b_I = (\boldsymbol{Z}' \boldsymbol{V}_{\bar{\boldsymbol{z}}_I}^{-1} \boldsymbol{Z})^{-1} \boldsymbol{Z}' \boldsymbol{V}_{\bar{\boldsymbol{z}}_I}^{-1} \bar{\boldsymbol{z}}_{\boldsymbol{I}}$ can be obtained via weighted least squares with consistent variance estimator $V_{b_I} = (\boldsymbol{Z}' \boldsymbol{V}_{\bar{\boldsymbol{z}}_I}{}^{-1} \boldsymbol{Z})^{-1}$. An associated

matched pairs odds ratio for the treatment effect is $b_I/(1-b_I)$, with $100^*(1-\alpha)\%$ confidence interval $(b_{I,L}/(1-b_{I,L}), b_{I,U}/(1-b_{I,U}))$ where $b_{I,L} = b_I - z_{1-\alpha/2} * \sqrt{V_{b_I}}$ and $b_{I,U} = b_I + z_{1-\alpha/2} * \sqrt{V_{b_I}}$.

Analysis may also proceed on $d_{h,I}$ for the informative pairs where $d_{h,I} = 1$ if Treatment 1 is a success and Treatment 2 is a failure, and $d_{h,I} = -1$ if Treatment 1 is a failure and Treatment 2 is a success. Then $\bar{d}_I$ represents the difference in the proportion of pairs with Success/Failure and the proportion of pairs with Failure/Success. With a covariate-adjusted counterpart b_I^* of the difference in proportions $\bar{d}_I$ and variance estimator $V_{b_I^*}$ using the $d_{h,I}$, the matched pairs odds ratio is $(b_I^* + 1)/(1 - b_I^*)$ with $100^*(1-\alpha)\%$ confidence interval $((b_{I,L}^* + 1)/(1 - b_{I,L}^*), (b_{I,U}^* + 1)/(1 - b_{I,U}^*))$. Exact analysis can proceed in this setting as previously described using the variance under the null hypothesis in (2.6) where $\boldsymbol{f}_{h,I} = (d_{h,I}, \boldsymbol{u}'_{h,I})'$.

$$\boldsymbol{V}_{\bar{\boldsymbol{f}}_{\boldsymbol{I},\boldsymbol{0}}} = \frac{1}{(q')^2} \sum_{h=1}^{q'} \boldsymbol{f}_{h,I} \boldsymbol{f}'_{h,I} \tag{2.6}$$

Alternatively, one can include all the pairs in the analysis by forming a_{h1} and a_{h2}, where $a_{h1} = 1$ if $d_h = 1$ (0 else), and $a_{h2} = 1$ if $d_h = -1$ (0 else). The vector $\boldsymbol{g}_h = (a_{h1}, a_{h2}, \boldsymbol{u}'_h)'$ can be created, and the associated mean vector $\bar{\boldsymbol{g}} = (\bar{a}_1, \bar{a}_2, \bar{\boldsymbol{u}}')'$ and estimated covariance matrix $\boldsymbol{V}_{\bar{\boldsymbol{g}}}$ can be formed in ways comparable to $\bar{\boldsymbol{f}}$ and $\boldsymbol{V}_{\bar{\boldsymbol{f}}}$. A transformation (via Taylor series linearization) may then be applied such that the model $\boldsymbol{Z} = (1\ \boldsymbol{0}'_P)'$ can be fit to $\tilde{\bar{\boldsymbol{g}}} = (log_e(\bar{a}_1/\bar{a}_2), \bar{\boldsymbol{u}}')'$ using the transformed covariance matrix $\boldsymbol{V}_{\tilde{\bar{\boldsymbol{g}}}} = \boldsymbol{A}\boldsymbol{V}_{\bar{\boldsymbol{g}}}\boldsymbol{A}'$, where

$$\boldsymbol{A} = \begin{bmatrix} \frac{1}{\bar{a}_1} & \frac{-1}{\bar{a}_2} & \boldsymbol{0}_{\boldsymbol{1}\times\boldsymbol{P}} \\ \boldsymbol{0}_{\boldsymbol{P}\times\boldsymbol{1}} & \boldsymbol{0}_{\boldsymbol{P}\times\boldsymbol{1}} & \boldsymbol{I}_{\boldsymbol{P}\times\boldsymbol{P}} \end{bmatrix}$$

and where $\boldsymbol{0}_{\boldsymbol{P}\times\boldsymbol{1}}$ is a matrix of zeroes and $\boldsymbol{I}_{\boldsymbol{P}\times\boldsymbol{P}}$ is a $P \times P$ identity matrix. Then

$\tilde{b} = (\boldsymbol{Z}'\boldsymbol{V}_{\tilde{\bar{\boldsymbol{g}}}}^{-1}\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{V}_{\tilde{\bar{\boldsymbol{g}}}}^{-1}\tilde{\bar{\boldsymbol{g}}}$ represents the adjusted estimate of the log of the matched pairs odds ratio and a consistent estimator for the variance of $\tilde{b}$ is $V_{\tilde{b}} = (\boldsymbol{Z}'\boldsymbol{V}_{\tilde{\bar{\boldsymbol{g}}}}^{-1}\boldsymbol{Z})^{-1}$. A large-sample $100^*(1-\alpha)\%$ confidence interval can be obtained for the log odds ratio as $\tilde{b} \pm z_{1-\alpha/2} * \sqrt{V_{\tilde{b}}}$. This approach retains the covariate information from non-informative pairs so that similar covariate distributions in the treatment groups holds on the basis of a valid randomization.

An alternative to Taylor series linearization for obtaining a confidence interval for the odds ratio is Fieller's theorem (Fieller, 1954). Since the matched pairs odds ratio can be expressed as the ratio of two means, $\bar{a}_1/\bar{a}_2$, this approach is applicable. Using the vector $\bar{\boldsymbol{g}}$ and its covariance matrix $\boldsymbol{V}_{\bar{\boldsymbol{g}}}$, one obtains the covariate-adjusted estimates $\tilde{\boldsymbol{b}} = (\tilde{b}_1, \tilde{b}_2)'$ and its associated covariance matrix $\boldsymbol{V}_{\tilde{\boldsymbol{b}}}$ through weighted least squares. One then forms the adjusted matched pairs odds ratio as $\tilde{b}_1/\tilde{b}_2$, and the confidence limits are the solutions x of a quadratic equation $ax^2 + bx + c = 0$ where a, b, and c are as follows:

$$\begin{aligned} a &= \tilde{b}_2^2 - z_{1-\alpha/2}^2 * (v_{\tilde{b}_2}) \\ b &= 2 * [z_{1-\alpha/2}^2 * (v_{\tilde{b}_1,\tilde{b}_2}) - \tilde{b}_1 * \tilde{b}_2] \\ c &= \tilde{b}_1^2 - z_{1-\alpha/2}^2 * (v_{\tilde{b}_1}) \end{aligned} \tag{2.7}$$

Here, $v_{\tilde{b}_2}$ and $v_{\tilde{b}_1}$ are estimated variances of $\tilde{b}_2$ and $\tilde{b}_1$, respectively, and $v_{\tilde{b}_1,\tilde{b}_2}$ is their estimated covariance obtained from $\boldsymbol{V}_{\tilde{\boldsymbol{b}}}$. The quantity $z_{1-\alpha/2}^2 = 3.84$ in the case of a 95% confidence interval.

2.2.3 M:1 Matched Sets with Randomization

The methodology in Section 2.2.1 may readily be extended to M:1 matched sets with randomization. These types of sets may arise in randomized clinical trials where one patient is randomized to control and M patients are randomized to the test treatment

within the matched set. In practice, M usually ranges between two and four.

Assume, without loss of generality, that we have M:1 matched sets where M patients are randomized to Treatment 1 and one patient is randomized to Treatment 2. Let $y_{hik} = 1$ denote a success for patient k on the ith treatment ($i = 1, 2$) from set h ($h = 1, \ldots, q$), and let $y_{hik} = 0$ denote a failure for patient k on the ith treatment from set h. Here, $k = 1, \ldots, M$ if $i = 1$, and $k = 1$ if $i = 2$. Let $\bar{y}_{h1\cdot} = \frac{1}{M}\sum_{k=1}^{M} y_{h1k}$ be the proportion of successes for the M patients on Treatment 1 in set h. Additionally, let $\boldsymbol{x}_{hik}$ be the vector of covariates corresponding to the ith treatment for patient k from set h, and $\bar{\boldsymbol{x}}_{h1\cdot} = \frac{1}{M}\sum_{k=1}^{M} \boldsymbol{x}_{h1k}$ is the mean vector of covariates for the M patients on Treatment 1 from set h. A vector of the differences in outcomes for the two treatments can be formed as $d_h = (\bar{y}_{h1\cdot} - y_{h21})$. With $\boldsymbol{u}_h = (\bar{\boldsymbol{x}}_{h1\cdot} - \boldsymbol{x}_{h21})$ denoting the difference between treatments for the means of the covariables within a set, a vector $\boldsymbol{f}_h = (d_h, \boldsymbol{u}_h')'$ can be formed for each set h. The mean vector $\bar{\boldsymbol{f}} = (\bar{d}, \bar{\boldsymbol{u}}')' = \frac{1}{q}\sum_{h=1}^{q} \boldsymbol{f}_h$ can be formed with an estimated covariance matrix $\boldsymbol{V}_{\bar{\boldsymbol{f}}}$ similar to (2.1). A covariate-adjusted estimate of the difference in proportions for the treatments can then be obtained in a similar manner using weighted least squares methodology to obtain b as in (2.2) and V_b as in (2.3). Random imbalances in the means of the covariables between the treatment groups can be assessed using the statistic in (2.5).

Methodology for obtaining an exact p-value for the treatment comparison would proceed in a similar manner to the methods described in Section 2.2.1 with the use of $\boldsymbol{V}_{\bar{\boldsymbol{f}},\mathbf{0}}$ instead of $\boldsymbol{V}_{\bar{\boldsymbol{f}}}$. Under the null hypothesis that the response of each of the $M + 1$ members of each set would be the same regardless of treatment assignment, $\boldsymbol{V}_{\bar{\boldsymbol{f}},\mathbf{0}}$ would take the form in (2.7). As previously stated, exact inference would then be formally restricted to the randomized study population.

$$\boldsymbol{V}_{\bar{\boldsymbol{f}},\mathbf{0}} = \frac{(M+1)}{M^2 q^2} \sum_{h=1}^{q} \sum_{i=1}^{2} \sum_{k=1}^{n_{hi}} (\boldsymbol{g}_{hik} - \bar{\boldsymbol{g}}_h)(\boldsymbol{g}_{hik} - \bar{\boldsymbol{g}}_h)' \tag{2.7}$$

Here, $n_{h1} = M$, $n_{h2} = 1$, and $\boldsymbol{g}_{hik} = (y_{hik}, \boldsymbol{x}'_{hik})'$ denotes the compound vector of the response and covariates for the kth individual on the ith treatment from the hth matched set. Additionally, the mean of this vector for the the hth matched set is given as $\bar{\boldsymbol{g}}_h = \frac{1}{M+1} \sum_{i=1}^{2} \sum_{k=1}^{n_{hi}} \boldsymbol{g}_{hik}$. The formula (2.7) takes into account the $M+1$ possible ways to assign M individuals to Treatment 1 and one individual to Treatment 2. When $M = 1$, the formula reduces to the form in (2.4).

A covariate-adjusted odds ratio may also be obtained for the M:1 setting. The procedure begins with an unadjusted estimate of the odds ratio such as the Mantel-Haenszel estimate (Breslow and Day, 1980), (Mantel and Haenszel, 1959) as in (2.8).

$$\hat{\psi}_{mh} = \frac{\sum_{m=0}^{M} m n_{m,0}}{\sum_{m=0}^{M} (M-m) n_{m,1}} \tag{2.8}$$

Here, $n_{m,1}$ is the number of matched sets with m successes among the M subjects on test treatment and one success on control, and $n_{m,0}$ is the number of matched sets with m successes on test treatment and zero successes on control. One could also define $a_{h1} = m$ if there are m successes on test treatment when there is a failure on control ($m = 0, \ldots, M$) for matched set h, and define $a_{h1} = 0$ if there is success on control. Similarly, define $a_{h2} = M - m$ if there are m successes on test treatment when there is a success on control, and define $a_{h2} = 0$ if there is failure on control. Then $\sum_{h=1}^{q} a_{h1}$ equals the numerator of $\hat{\psi}_{mh}$ and $\sum_{h=1}^{q} a_{h2}$ equals the denominator of $\hat{\psi}_{mh}$. A vector $\boldsymbol{g}_h = (a_{h1}, a_{h2}, \boldsymbol{u}'_h)'$ can be formed. Its mean vector $\bar{\boldsymbol{g}}$ and covariance matrix $\boldsymbol{V}_{\bar{\boldsymbol{g}}}$ are formed as in Section 2.2.1. Covariate adjustment may then proceed through the application of weighted least squares to the transformed vector $\tilde{\bar{\boldsymbol{g}}} = (log_e(\bar{a}_1/\bar{a}_2), \bar{\boldsymbol{u}}')' = (log_e(\hat{\psi}_{mh}), \bar{\boldsymbol{u}}')'$ using its transformed covariance matrix. Construction of a large-sample confidence interval for the log odds ratio could proceed based on the covariate-adjusted estimate $\tilde{b}$ and its estimated variance $v_{\tilde{b}}$ as in Section 2.2.1. Alternatively, the covariate-adjusted estimate $\tilde{\boldsymbol{b}} = (\tilde{b}_1, \tilde{b}_2)'$ and its variance matrix $\boldsymbol{V}_{\tilde{\boldsymbol{b}}}$ could be estimated from

weighted least squares using $\bar{\boldsymbol{g}}$ and $\boldsymbol{V}_{\bar{\boldsymbol{g}}}$, and Fieller's theorem is then applied to create a confidence interval for the ratio $\tilde{b}_1/\tilde{b}_2$.

2.2.4 1:1 Matched Case-Control Studies

Matched sets often arise in non-randomized case-control studies where interest lies in assessing the relationship between a dichotomous exposure (Yes/No) and case-control status. The mechanism of NPANCOVA can be applied to a 1:1 matched case-control study as follows: Define $y_{h1} = 1$ if the case from pair h was exposed and $y_{h1} = 0$ if that case was not exposed. Similarly, define $y_{h2} = 1$ if the control from pair h was exposed and $y_{h2} = 0$ if not exposed. Let $\boldsymbol{x}_{h1}$ be the covariates for a case, and let $\boldsymbol{x}_{h2}$ be the covariates for a control. It should be noted that in the case-control setting, the difference in the means of covariables defined by case-control status is not likely zero unless the covariables are not predictive of case-control status.

Consider $d_h = y_{h1}(1 - y_{h2}) - y_{h2}(1 - y_{h1})$. Then $d_h = 1$ if the case from pair h is exposed and the control from pair h is not, $d_h = 0$ if the case and control have the same exposure status, and $d_h = -1$ if the control is exposed and the case is not. Further define $\boldsymbol{c}_h = y_{h1}(1 - y_{h2})(\boldsymbol{x}_{h1} - \boldsymbol{x}_{h2}) + y_{h2}(1 - y_{h1})(\boldsymbol{x}_{h2} - \boldsymbol{x}_{h1})$. Note that $\boldsymbol{c}_h = d_h(\boldsymbol{x}_{h1} - \boldsymbol{x}_{h2})$. Thus, when the case is exposed and the control is not exposed in pair h, $\boldsymbol{c}_h$ represents the difference in the covariables of exposed and unexposed. When the control is exposed and the case is not exposed in pair h, $\boldsymbol{c}_h$ represents the difference in the covariables of exposed and unexposed as well.

The NPANCOVA methodology may then be applied to a vector $\boldsymbol{f}_h = (d_h, \boldsymbol{c}_h')'$ as in the setting with randomization. Use of $\boldsymbol{c}_h$ ensures the difference in the means of covariables will be defined by exposure status rather than case-control status, although the analysis excludes sets where both the case and control within a pair are exposed or both the case and control are not exposed (since $\boldsymbol{f}_h$ would be a vector of zeros by

definition). The odds ratio and its confidence interval can be formed via the methodology relating to the informative pairs analysis on the $d_{h,I}$ in Section 2.2.2, and pairs are excluded when the $d_h = 0$. The mean vector $\bar{\boldsymbol{f}}_I$ is formed as before and $\boldsymbol{V}_{\bar{\boldsymbol{f}}_I}$ is created as in (2.1). The covariate-adjusted estimate b_I^* is formed as in (2.2) and its consistent variance estimator $V_{b_I^*}$ is formed as in (2.3). It should be noted that while this odds ratio is technically the exposure odds ratio, this quantity is algebraically equivalent to the disease odds ratio of interest. Assessment of goodness-of-fit (in the sense of comparable distributions of covariates for the exposure groups) may proceed using the statistic in (2.5) using data for the informative pairs analysis.

2.3 Examples

2.3.1 Multi-Center 1:1 Clinical Trial Data for Improvement of a Skin Condition

Researchers enrolled one pair of individuals from each of 79 randomly selected clinics. For each pair, one individual was randomly chosen to receive the test treatment, and another individual was randomly chosen to receive placebo. The initial grade of a skin condition (coded 1-4 for mild to severe) was recorded for each patient. The outcome of improvement in the skin condition was recorded as a 1 for improvement and 0 for no improvement. A matched pairs analysis is appropriate for these data, with adjustment for initial grade of the skin condition. Thirty-four (43%) pairs showed improvement for the patient receiving test treatment and no improvement for the patient receiving placebo. Twenty (25%) pairs showed improvement for the placebo patient but not the patient on test treatment. The remaining 25 (32%) pairs were concordant (both members of a pair showed improvement or both showed no improvement). The data are available in (Stokes, Davis and Koch, 2012).

Results are presented in Table 2.1. With no covariate adjustment, the difference

in proportions of individuals with improvement on test (but not placebo) and placebo (but not test) was $d = ((34/79) - (20/79)) = 0.1772$ (S.E.=$\sqrt{v_d}$=0.0909). With respect to $\sqrt{v_{d,0}}$ under the null hypothesis, the unadjusted McNemar's test statistic $d^2/v_{d,0}$ had p=0.0567 from its approximate chi-squared distribution with d.f.=1; and its exact counterpart from McNemar's test was 0.0759 via the EXACT MCNEM option in SAS PROC FREQ. When covariate adjustment was made, the adjusted estimate of the difference in proportions was 0.1930 with an associated 13% reduction in the standard error (S.E.=0.0791). The treatment effect was found to be statistically significant via $Q_{b,0} = b_0^2/v_{b,0}$ with p=0.0181; the corresponding exact counterpart had p=0.0179 based on $N = 100,000$ re-randomizations for the NPANCOVA analysis. The approximate two-sided 95% confidence interval for the difference in proportions from NPANCOVA was (0.0380, 0.3480). An approximation using the t-distribution with $(q - q_0 - 1 - P) = (79 - 25 - 1 - 1) = 52$ degrees of freeom was (0.0343, 0.3517). No evidence of a random imbalance in initial skin condition between the two treatment groups was found ($Q_{\bar{\boldsymbol{u}}} = 0.1181$, df=1, p=0.7311).

Table 2.1: Matched Pairs Analysis of Multi-Center Clinical Trial Data from Example 2.3.1

Method	Estimate	S.E.	OR	95% CI		p-value	Exact p-value
				Lower	Upper		
Difference of Proportions- All Pairs							
Unadjusted ($q = 79$)	0.1772	0.0909	–	-0.0009	0.3553	0.0567[b]	0.0759[b]
NPANCOVA[a] ($q = 79$)	0.1930	0.0791	–	0.0380	0.3480	0.0181[b]	0.0179[b]
Proportion of Yes/No - Informative Pairs							
Unadjusted on $a_{h,I}$ ($q' = 54$)	0.6296	0.0663	1.7000	0.9986	3.1588	0.0567[b]	0.0759[b]
NPANCOVA[a] **on** $a_{h,I}$ ($q' = 54$)	0.6262	0.0556	1.6754	1.0713	2.7770	0.0288[b]	–
Difference of Proportions - Informative Pairs							
Unadjusted on $d_{h,I}$ ($q' = 54$)	0.2593	0.1327	1.7000	0.9986	3.1588	0.0567[b]	0.0759[b]
NPANCOVA[a] **on** $d_{h,I}$ ($q' = 54$)	0.2525	0.1112	1.6754	1.0713	2.7770	0.0288[b]	0.0280[b]
Log Odds Ratio - Informative Pairs ($q' = 54$)							
Unadjusted Conditional Logistic Regression	0.5306	0.2818	1.7000	0.9785	2.9533	0.0567[c]	0.0759[c]
Conditional Logistic Regression[a]	0.7113	0.3487	2.0366	1.0282	4.0340	0.0353[c]	0.0377[c]
Log Odds Ratio - All Pairs							
NPANCOVA[a] ($q = 79$)	0.5790	0.2462	1.7842	1.1012	2.8909	0.0187[d]	–

[a] Adjusted for initial skin condition

[b] Unadjusted (i.e. McNemar) and NPANCOVA p-values based on test using randomization-based covariance matrix in (2.4)

[c] P-values from score test of coefficient for treatment in conditional logistic regression

[d] NPANCOVA p-value based on test using covariance matrix in (2.1)

When only informative strata were considered in the analysis, the adjusted proportion of individuals with a success on test treatment (and failure on placebo) was 0.6262. This was statistically different from 0.5 with a p-value of 0.0288. The odds of success on test treatment (and failure on placebo) were (0.6262/0.3738)=1.6754 times the odds of success on placebo (and failure on test). When considering the difference in proportions for the informative pairs, the resulting odds ratio and p-value were identical, and the exact p-value was 0.0280. Among the informative strata, there was no evidence of a random imbalance in initial skin condition between the two treatment groups ($Q_{\bar{\boldsymbol{u}}} = 0.0088$, df=1, p=0.9251).

The unadjusted McNemar's odds ratio was $34/20 = 1.7000$, whereas when adjusting for intial skin condition, conditional logistic regression produced an odds ratio of 2.0366, and this estimate was statistically significant (p=0.0353). The NPANCOVA analysis on all pairs produced an adjusted odds ratio (1.7842) similar to the McNemar's odds ratio, and a 13% reduction in the standard error for its logarithm (from the McNemar's S.E. of 0.2818) was observed (S.E.=0.2462). This resulted in a statistically significant p-value for the treatment effect (p=0.0187). It should be noted that the estimated odds ratio for NPANCOVA tends to be closer to the null of 1.0 than the conditional logistic regression estimate because the NPANCOVA estimate is like a population-averaged estimate relative to covariates for matched subjects, and the conditional logistic regression estimate has a subject-specific interpretation for matched patients with the same initial skin condition.

The 95% large-sample confidence interval for the McNemar's odds ratio was (0.9785, 2.9533), while a slightly more conservative interval using Fieller's theorem (data not shown) produced an interval of (0.9941, 3.1794). For the covariate-adjusted odds ratio, the large-sample interval was (1.1012, 2.8909), and the interval based on Fieller's

theorem was (1.1198, 3.0913). While the large-sample intervals were based on estimates where the log odds ratio was first created then underwent covariate adjustment, the Fieller's theorem intervals were based on creating a ratio of the covariate-adjusted means. In the unadjusted case, the estimated matched pairs odds ratio was 1.7 regardless of method. For the adjusted case, the different approaches produced similar estimated odds ratios (1.7842 for NPANCOVA, 1.7846 for ratio of adjusted means).

2.3.2 Extension of Multi-Center Clinical Trial Data to 2:1 Setting

To illustrate the application of the NPANCOVA methodology to a randomized clinical trial with 2:1 matching, the data from Example 2.3.1 were re-examined. The 54 informative pairs and the 25 non-informative pairs were ordered separately by the average baseline age for a pair. Starting with the smallest average baseline age, the first non-informative pair was divided so that the patient on treatment was assigned to treatment for the first informative pair, and the patient on placebo was assigned to placebo for the second informative pair. The next non-informative pair on the list was divided so that the patient on placebo was assigned to placebo for the third informative pair, and the patient on treatment was assigned to treatment for the fourth informative pair. This procedure continued (with the division of the 53rd informative pair and deletion of the 54th informative pair) until 26 2:1 sets and 26 1:2 sets were created. Since age had essentially no association with the response (within pairs), this assignment was essentially random and kept all but one of the original pairs in the analysis data set.

Since the data contain matched sets of different type (e.g. 2:1 and 1:2), a stratified application of NPANCOVA was warranted (Koch et al., 1998). The vectors $\bar{d}_1$ and $\bar{\boldsymbol{u}}_1$ were formed for the 2:1 sets, and $\bar{d}_2$ and $\bar{\boldsymbol{u}}_2$ were formed for the 1:2 sets as in Section

2.2.1. Separate $\bar{\boldsymbol{f}}_1$ and $\bar{\boldsymbol{f}}_2$ vectors and covariance matrices $\boldsymbol{V}_{\bar{\boldsymbol{f}}_1}$ and $\boldsymbol{V}_{\bar{\boldsymbol{f}}_2}$ were also formed as appropriate. A weighted estimate $\bar{\boldsymbol{f}}_{\boldsymbol{w}} = (q_1\bar{\boldsymbol{f}}_1 + q_2\bar{\boldsymbol{f}}_2)/(q_1+q_2)$ and weighted covariance matrix $\boldsymbol{V}_{\bar{\boldsymbol{f}}_{\boldsymbol{w}}} = (q_1^2\boldsymbol{V}_{\bar{\boldsymbol{f}}_1} + q_2^2\boldsymbol{V}_{\bar{\boldsymbol{f}}_2})/(q_1+q_2)^2$ were created, where $q_1 = 26$ and $q_2 = 26$ were the numbers of 2:1 and 1:2 sets, respectively. Weighted least squares was then applied to $\bar{\boldsymbol{f}}_{\boldsymbol{w}}$ using $\boldsymbol{V}_{\bar{\boldsymbol{f}}_{\boldsymbol{w}}}$ to obtain an estimate adjusted for initial skin condition. If the number of matched sets for each type of allocation was ≥ 50, covariate adjustment could have been applied to the 2:1 sets and 1:2 sets separately before weighting the estimates across type of matched set.

A stratified version of the methodology for the odds ratio can be applied in a similar manner. For the 2:1 sets, the procedure for obtaining the estimate and its covariance matrix is as described in Section 2.2.3. For the 1:2 sets, $a_{h1} = 2, 1$, or 0 if 0,1, or 2 placebo successes were observed in set h when a test success was observed and is 0 if a test failure was observed, and $a_{h2} = 0, 1$, or 2 if 0,1, or 2 placebo successes were observed in set h when a failure on test was observed and is 0 if a test success was observed. A vector $\bar{\boldsymbol{g}}_j = (\bar{a}_{1j}, \bar{a}_{2j}, \bar{\boldsymbol{u}}_j')'$ was formed for each type of allocation so that $j = 1$ for the 2:1 sets and $j = 2$ for the 1:2 sets. Corresponding covariance matrices $\boldsymbol{V}_{\bar{\boldsymbol{g}}_{\boldsymbol{j}}}$ $(j = 1, 2)$ can be formed through the expression in (2.1). A weighted vector $\bar{\boldsymbol{g}}_w = (\bar{a}_{1w}, \bar{a}_{2w}, \bar{\boldsymbol{u}}_w')'$ and its covariance matrix $\boldsymbol{V}_{\bar{\boldsymbol{g}}_{\boldsymbol{w}}}$ are then formed as above. Then the Mantel-Haenszel estimate of the log of the odds ratio for test vs. placebo would be given as $log_e(\bar{a}_{1w}/\bar{a}_{2w})$ and covariate adjustment could proceed via weighted least squares with the vector $\tilde{\bar{\boldsymbol{g}}}_w = (log_e(\bar{a}_{1w}/\bar{a}_{2w}), \bar{\boldsymbol{u}}_w')'$, covariance matrix $\boldsymbol{V}_{\tilde{\bar{\boldsymbol{g}}}_{\boldsymbol{w}}} = \boldsymbol{A}\boldsymbol{V}_{\bar{\boldsymbol{g}}_{\boldsymbol{w}}}\boldsymbol{A}'$, and model $\boldsymbol{Z} = (1\ \boldsymbol{0}_P')'$.

Results are presented in Table 2.2. The unadjusted difference in proportions between the test and placebo groups was 0.1923 (S.E.=0.1084). At the 0.05 level, this result was not significantly different from zero with p=0.0761. However, with respect to the null variance from (2.7) with weighting across the strata (S.E.=0.0981), p=0.0499,

with this being identical to the Mantel-Haenszel p-value from SAS PROC FREQ. When adjusting for initial skin condition, the difference in proportions for NPANCOVA using the variance form in (2.1) weighted across the strata was 0.2087 (S.E.=0.0865). This was statistically significant with a p-value of 0.0158. The assessment for random imbalances in mean initial skin condition between the two treatment groups had $Q_{\bar{\boldsymbol{u}}} = 0.0632$ with df=1 and p=0.8015, indicating no evidence of an imbalanced randomization with respect to initial skin condition.

Table 2.2: Analysis of Multi-Center 2:1 and 1:2 Clinical Trial Data from Example 2.3.2

Method	Estimate	S.E.	OR	95% CI		p-value	Exact p-value
				Lower	Upper		
Difference in Proportions							
Unadjusted	0.1923	0.1084	–	-0.0202	0.4048	0.0499[b]	0.0563[b]
NPANCOVA[a]	0.2087	0.0865	–	0.0392	0.3783	0.0158[c]	–
Log Odds Ratio							
Unadjusted	0.5247	0.3093	1.6897	0.9215	3.0980	0.0899[c]	–
Unadjusted Conditional Logistic Regression	0.5903	0.3040	1.8045	0.9945	3.2744	0.0499[d]	0.0560[d]
NPANCOVA[a]	0.5711	0.2478	1.7701	1.0891	2.8770	0.0212[c]	–
Conditional Logistic Regression[a]	0.7245	0.3380	2.0637	1.0640	4.0027	0.0287[d]	0.0385[d]

[a] Adjusted for initial skin condition

[b] P-value based on test using randomization-based covariance matrix in (2.7)

[c] NPANCOVA p-values based on test using covariance matrix as in (2.1)

[d] P-values from score test of coefficient for treatment in conditional logistic regression

1 Test : 2 Placebo

		Placebo		
		0	1	2
Test	1	4	14	0
	0	0	3	5

2 Test : 1 Placebo

		Placebo	
		0	1
	2	12	0
Test	1	3	6
	0	0	5

Figure 2.1: Frequencies of Success Outcomes for Treatment (T) * Placebo (P) for 1:2 and 2:1 Matched Sets for Data from Example 2.3.2

Regarding the odds ratio, consider the data in Figure 2.1. When considering the 1:2 matched sets, the unadjusted M-H estimate of the odds ratio is $\hat{\psi}_{mh} = (2(4) + 1(14) + 0(0))/(0(0) + 1(3) + 2(5)) = 1.6923$. For the 2:1 matched sets, the unadjusted M-H estimate of the odds ratio is $\hat{\psi}_{mh} = (0(0) + 1(3) + 2(12))/(2(5) + 1(6) + 0(0)) = 1.6875$. Table 2.2 contains the estimate of the odds ratio after combining the 1:2 and 2:1 estimates. The unadjusted odds ratio is the ratio of the sum of the numerators in the 1:2 and 2:1 estimates divided by the sum of the denominators and was $(8 + 14 + 3 + 24)/(3 + 10 + 10 + 6) = 49/29 = 1.6897$.

The NPANCOVA analysis produced an odds ratio of 1.7701, indicating that the odds of improvement on test treatment are roughly 1.8 times the odds of improvement on placebo after adjusting for initial skin condition. This odds ratio is closer to the null than the odds ratio obtained from conditional logistic regression (2.0637), which is expected given the NPANCOVA odds ratio is a population-averaged estimate (relative to covariates) for matched subjects; the odds ratio from conditional logistic regression is subject-specific with respect to matched patients with the same initial skin condition. A 20% reduction in the standard error for the log odds ratio was observed (S.E.=0.2478) as compared to the unadjusted analysis (S.E.=0.3093). The p-value for the unadjusted analysis was not significant (p=0.0899) at the 0.05 level, but the p-values for the NPANCOVA analysis (p=0.0212) and conditional logistic regression

(p=0.0287) were statistically significant.

A 95% confidence interval for the unadjusted odds ratio was (0.9215, 3.0980), and an interval based on Fieller's theorem (data not shown) was (0.9475, 3.3624). The estimated odds ratio for the unadjusted case is 1.6897, and in the case of Fieller's, the ratio of unadjusted means was also estimated as 1.6897. After covariate adjustment, the large-sample confidence interval was (1.0891, 2.8770) with an odds ratio estimate of 1.7701. Using Fieller's theorem, the confidence interval was (1.1103, 3.0629) with the ratio of covariate-adjusted means being 1.7706.

2.3.3 1:1 Matched Case-Control Study of Vaccine Exposure and Influenza

Researchers tracked cases of influenza requiring hospitalization in residents of a mid-western county, aged 65 and older during a two-month period in one winter. Each case was sex- and age-matched to two controls, and information was obtained on whether study participants had a vaccine shot and whether they had lung disease. The data for these 150 matched sets are available in (Stokes, Davis and Koch, 2012). For illustration, the data from each case was duplicated and then separately assigned to each of the two controls within a set to create 300 case-control pairs matched on the basis of sex and age.

Results are presented in Table 2.3. When considering only the informative pairs (145 matched pairs), the unadjusted analysis produced a log odds ratio of -0.3199 (S.E.=0.1688). This was not statistically significant at the 0.05 level (p=0.0580). A 4.9% reduction in the standard error was observed with the NPANCOVA adjusted analysis of the log odds ratio (S.E.=0.1605). The odds ratio for NPANCOVA was 0.7214, and the p-value was p=0.0420. The conditional logistic regression analysis produced an adjusted log odds ratio of -0.3616 (S.E.=0.1781). Thus, the odds of being a case for

vaccinated individuals compared to unvaccinated individuals was exp(-0.3616)=0.6966, after adjusting for lung disease and matched set. Statistical significance was similarly observed at the 0.05 level with p=0.0409. For NPANCOVA, the test of goodness-of-fit (in the sense of similar proportions of patients with lung disease in the vaccinated and unvaccinated groups) was not contradictory ($Q_{\bar{\boldsymbol{u}}} = 0.0158$, df=1, p=0.9001). As in Example 2.3.1, the adjusted odds ratio for NPANCOVA tends to be closer to the null value of 1.0 than the estimate from conditional logistic regression since it has an interpretation as a population-averaged estimate (relative to covariates) for matched subjects, and the estimate from conditional logistic regression has an interpretation as an estimate which is applied to matched subjects with the same status for lung disease.

Table 2.3: Analysis of 1:1 Matched Case-Control Study From Example 2.3.3

Method	Est.	S.E.	OR	95% CI		p-value	Exact p-value
				Lower	Upper		
Log Odds Ratio - Informative Pairs							
Unadjusted	-0.3199	0.1688	0.7262	0.5216	1.0110	0.0580[b]	0.0673[c]
Conditional Logistic Regression[a]	-0.3616	0.1781	0.6966	0.4913	0.9876	0.0409[c]	0.0454[c]
NPANCOVA[a]	-0.3265	0.1605	0.7214	0.5267	0.9882	0.0420[b]	–

Exposure is vaccination (Yes/No)

[a] Adjusted for lung disease (Yes/No)

[b] Unadjusted and NPANCOVA p-values based on test using covariance matrix as in (2.1)

[c] P-value from score test of coefficient for vaccination in conditional logistic regression

2.4 Discussion

This chapter outlines methodology for applying nonparametric randomization-based analysis of covariance to dichotomous outcomes from matched sets with very small sample sizes within the matched sets. The methods presented here are an alternative to McNemar's test in the case of matched pairs or conditional logistic regression when covariate adjustment is desired. NPANCOVA only requires a valid randomization to the two treatment groups as well as large enough sample size so that weighted least squares estimates have approximate multivariate normal distributions. However, exact inference for testing null hypotheses may be performed for the permutation distributions of the weighted least squares estimates. In the case of an observational matched case-control study, the NPANCOVA methodology is applicable to informative pairs provided the distributions of the covariables are similar by exposure status. For the 1:1 randomized trial and 2:1 randomized trial examples, variance reduction was obtained after covariance adjustment with NPANCOVA (as compared to the unadjusted analyses). In the 1:1 matched case-control study, variance reduction was observed as compared to the unadjusted analysis.

The NPANCOVA methods presented here are appropriate for the regulatory setting where statistical assumptions of parametric modeling methods can not be verified prior to the collection of data. Conditional logistic regression methods for matched sets make assumptions about the relationship between the log odds of response and covariates. Additionally, standard errors for the treatment effect in a conditional logistic regression model may not necessarily decrease with covariance adjustment. It has been observed that the increase in standard error may be accompanied by an increase in the estimated treatment effect, which may provide a p-value that is comparable to the p-value obtained from NPANCOVA (Tangen and Koch, 1999*a*). The NPANCOVA methods presented here avoid any modeling assumptions about the relationship between the

outcomes and covariables for adjustment, and standard errors for the treatment effect estimates do not increase as a consequence of adjustment.

The methodology which restricts analysis to only informative pairs makes an assumption that the distributions of the covariates remain similar between treatment groups after the removal of non-informative pairs. In the case of the 1:1 matched case-control study, the assumption is that the exposure groups have similar distributions of adjustment covariables. The estimated treatment effect may be subject to bias if this assumption does not hold, as forcing the differences in means of covariables to zero may no longer be appropriate. Assessment of goodness-of-fit of the weighted least squares model can be done through the statistic in (2.5). In the 1:1 randomized case, we offer the methodology which includes all pairs as a way of avoiding this assumption about non-informative pairs. For the matched case-control study, the assumption may be reasonable if the influence of unobserved covariates is thought to be negligible after matching and covariate adjustment for potential confounders.

The methods presented in this chapter are intended for application to matched pairs and matched sets of size ≤ 10. For studies with matched sets of larger size, methodology in the next chapter may be more easily implemented. The methodology involves using the DFBETA diagnostics from a treatment-only conditional logistic regression model to estimate the covariance of the unadjusted log odds ratio. Ths covariance estimate can then be combined with information relating to the covariance matrix for the differences in means of covariables in order to produce a covariate-adjusted estimate of the log odds ratio via weighted least squares. This approach is also extended to time-to-event outcomes in the presence of stratification.

Chapter 3

Nonparametric Randomization-Based Covariate Adjustment for Stratified Analysis of Time-to-Event or Dichotomous Outcomes

3.1 Introduction

For studies which have matched sets and dichotomous outcomes, conditional logistic regression is an analysis method which is often used in order to estimate odds ratios while adjusting for other covariables. Adjusting for the strata formed by the matched sets could be done through indicator variables in an unconditional logistic regression, but bias in model fitting can become an issue if the number of terms in the model is large relative to the available sample size, or the sample size per stratum is small (e.g. < 10) for some strata (Breslow and Day, 1980). Thus, conditional logistic regression is a method which removes the variability among the strata and results in estimation of relationships for explanatory variables with variation within strata. Maximum likelihood estimates of odds ratios can be obtained via maximizing a conditional likelihood. This conditional likelihood is formed by conditioning on sufficient statistics for the

strata-specific intercepts in the model. This results in a regression model which eliminates those extra parameters and provides an estimated effect for treatment or other covariates. However, the use of a conditional likelihood means no individual stratum effects can be estimated; they are considered nuisance parameters.

Stratification may also be accommodated in studies where the outcome is a time-to-event outcome. The Cox proportional hazards model (Cox, 1972) is frequently used for analyzing time-to-event outcomes in the presence of covariables. This is a model for the hazard of the event, and it consists of the product of two parts. The first relates to the underlying baseline hazard, which is not assumed to have any particular functional form. The second part relates to the covariables, and it consists of the exponentiation of a linear combination of P covariables with parameters $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_P)'$. The net effect of a one-unit increase in the jth covariate corresponds to a multiplicative increase of $exp(\beta_j)$ in the hazard of the event. A stratified analysis may be conducted with the Cox model by assuming separate baseline hazard functions for each level of a stratification factor. This type of analysis may be reasonable if factors such as center (in a multicenter clinical trial) or geographic region are thought to have heterogeneity in the underlying baseline hazard with estimation of hazard ratios for these factors being not of interest.

For both the conditional logistic regression model for dichotomous outcomes and the proportional hazards model for time-to-event outcomes, regression diagnostics can be used to assess the fit of the model. One diagnostic, the DFBETA residual, is calculated for each regression parameter for each independent observation in the data set. Its value corresponds to the absolute change in the estimated regression parameter for the full data set and the estimated regression parameter for the data set with that observation omitted. For most applications, including the ones presented in this chapter, the DFBETA residual is approximated via a one-step procedure. This involves

starting at the estimated regression parameter for the full data set, deleting an observation, and making one more Newton-Raphson step to obtain the estimated regression parameter with that observation omitted (Pregibon, 1981). Large values of DFBETA imply stronger influence of a particular observation on the parameter estimates. These diagnostics have been developed for unconditional logistic regression (Pregibon, 1981), conditional logistic regression (Storer and Crowley, 1985), proportional hazards models (Cain and Lange, 1984), and for generalized estimating equations (Preisser and Qaqish, 1996). Wei *et. al.* developed methodology for a robust estimate of the covariance matrix of $\hat{\boldsymbol{\beta}}$ by using the sum (over the observations) of the cross-products of DFBETA residuals (Wei, Lin and Weissfeld, 1989). This approach also applies to multivariate time-to-event outcomes, and the DFBETA residuals can easily be obtained from the PHREG procedure in SAS software.

In Section 1.5, methodology was described for applying nonparametric covariance adjustment to an unadjusted log hazard ratio in a randomized clinical trial with a time-to-event outcome and two treatment groups (Saville and Koch, 2013). A Cox proportional hazards model was fit with a single indicator for treatment, and the associated log hazard ratio β was estimated. Additionally, through the methodology of Wei *et. al.*, the sum of the squared DFBETAs from this model were used to obtain a robust covariance estimate $v_{\hat{\beta}}$ (Wei, Lin and Weissfeld, 1989). Once these quantities were obtained, they were combined with covariate information, and weighted least squares was then used to adjust $\hat{\beta}$ for the covariates.

This chapter extends the methodology of Saville and Koch (2013) to situations involving stratification. In the case of a dichotomous outcome with matched sets, the DFBETA residuals are obtained from a treatment-only conditional logistic regression model. In the case of a stratified time-to-event analysis, the DFBETA residuals are

obtained from a treatment-only Cox proportional hazards model which addresses possibly different baseline hazard functions for each stratum. Covariance structure for the stratification adjusted log odds ratio (or hazard ratio) is estimated through the methods of Wei *et. al.*. Covariance matrices for the covariates are derived through methods from Saville and Koch (2013) which produce a covariance matrix analogous to one presented in Koch *et. al.* (1998). Estimates of the corresponding covariances between the respective log hazard ratio(s) and the respective differences in means of covariables are provided through the sums of the cross-products of the DFBETA residuals with the quantities derived from averages of pairwise differences in covariables for the respective individuals versus all individuals in the opposite treatment group. The appendices present other specifications of the covariance structure derived from multivariate U-statistics methods. The information from the treatment-only model and the covariance matrix for the covariates is combined to provide nonparametric covariance adjustment of the log odds ratio (or log hazard ratio) via weighted least squares. This approach provides covariate adjustment which relies only on a valid stratified randomization and does not make any assumptions about the relationships between the covariables and the outcome. Methodology is first presented separately for the log odds ratio and the log hazard ratio followed by examples and discussion.

3.2 Methods

3.2.1 Hybrid Methodology for Dichotomous Outcomes with Stratification

Let π_{hj} denote the probability of an outcome of interest for the jth individual ($j = 1, \ldots, n_h$) in the hth stratum ($h = 1, \ldots, q$) who is receiving either the test treatment or the control. Here, $n = \sum_{h=1}^{q} n_h$ represents the total number of individuals being studied. Assume that individuals on the study were randomized within strata

to either test treatment or a comparator (which may or may not be placebo). Let $x_{hj} = 1$ if test treatment was received and $x_{hj} = 0$ if comparator was received. Then, a conditional logistic regression model only including treatment can be written as

$$\log(\frac{\pi_{hj}}{1-\pi_{hj}}) = \alpha_h + \beta x_{hj} \tag{3.1}$$

where the α_h is a stratum-specific intercept, and β represents the log of the odds ratio comparing the odds of the outcome for test treatment to the odds of the outcome on comparator within a stratum. This model eliminates the variability among strata, focusing instead on assessment of a treatment effect within strata.

Let $i = 1, 2$ index the test and comparator groups, respectively, and let $\boldsymbol{r}_{hi} = (r_{hi1}, \ldots, r_{hin_{hi}})'$ be the $n_{hi} \times 1$ vector of unstandardized DFBETA residuals for the n_{hi} individuals on treatment i in stratum h so that $\boldsymbol{r} = (\boldsymbol{r}'_{11}, \boldsymbol{r}'_{12}, \ldots, \boldsymbol{r}'_{q1}, \boldsymbol{r}'_{q2})'$ is the $n \times 1$ vector of DFBETA residuals from fitting the treatment-only conditional logistic regression model in (3.1) to all $n = \sum_{h=1}^{q} \sum_{i=1}^{2} n_{hi}$ individuals in the trial. Furthermore, let $n_h = n_{h1} + n_{h2}$ for stratum h and let $W = \sum_{h=1}^{q} (n_{h1} n_{h2})/n_h$. Also define $\boldsymbol{x}_{hik}$ as the $P \times 1$ vector of covariates for the kth individual on treatment i from stratum h, and let $\boldsymbol{X}'_{hi} = (\boldsymbol{x}_{hi1}, \ldots, \boldsymbol{x}_{hin_{hi}})$ be a $P \times n_{hi}$ matrix of the covariates for the individuals on treatment i in stratum h. With $\bar{\boldsymbol{x}}_{hi} = \sum_{k=1}^{n_{hi}} \boldsymbol{x}_{hik}/n_{hi}$ representing the sample mean vector of the covariates for treatment i from stratum h, the $n_{hi} \times P$ matrix $C_{hi} = \frac{n_{h1} n_{h2}}{n_h W \sqrt{n_{hi}(n_{hi}-1)}} (\boldsymbol{X}_{hi} - \mathbf{1}\bar{\boldsymbol{x}}'_{hi})$ can be created where $\mathbf{1}$ is a $n_{hi} \times 1$ vector of ones. An unbiased estimate for the covariance matrix for the covariate means for treatment i in stratum h is in (3.2).

$$\begin{aligned} \boldsymbol{V}_{\bar{\boldsymbol{x}}_{hi}} &= \left(\frac{n_{h1} n_{h2}}{n_h W}\right)^2 \frac{(\boldsymbol{X}_{hi} - \mathbf{1}\bar{\boldsymbol{x}}'_{hi})'(\boldsymbol{X}_{hi} - \mathbf{1}\bar{\boldsymbol{x}}'_{hi})}{n_{hi}(n_{hi}-1)} \\ &= \boldsymbol{C}'_{hi} \boldsymbol{C}_{hi} \end{aligned} \tag{3.2}$$

With the $(P+1)\times 1$ vector $\boldsymbol{f} = (\hat{\beta}, \bar{\boldsymbol{g}}')' = (\hat{\beta}, (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)')'$ containing the estimated log odds ratio $\hat{\beta}$ from the treatment-only conditional logistic regression model and the weighted difference in means of the covariates $\bar{\boldsymbol{g}} = (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)$, where $\bar{\boldsymbol{x}}_i = \sum_{h=1}^{q}(n_{h1}n_{h2}\bar{\boldsymbol{x}}_{hi})/W$, a consistent estimator for the covariance matrix of $\boldsymbol{f}$ is provided by $\boldsymbol{V_f}$ in (3.3).

$$\begin{aligned}\boldsymbol{V_f} &= \begin{bmatrix} \boldsymbol{r}'\boldsymbol{r} & \sum_{h=1}^{q}(\boldsymbol{r}'_{h1}\boldsymbol{C}_{h1} - \boldsymbol{r}'_{h2}\boldsymbol{C}_{h2}) \\ \sum_{h=1}^{q}(\boldsymbol{C}'_{h1}\boldsymbol{r}_{h1} - \boldsymbol{C}'_{h2}\boldsymbol{r}_{h2}) & \sum_{h=1}^{q}\sum_{i=1}^{2}\boldsymbol{C}'_{hi}\boldsymbol{C}_{hi} \end{bmatrix} \\ &= \begin{bmatrix} v_{\hat{\beta}} & \boldsymbol{V}'_{\hat{\beta},\bar{\boldsymbol{g}}} \\ \boldsymbol{V}_{\hat{\beta},\bar{\boldsymbol{g}}} & \boldsymbol{V}_{\bar{\boldsymbol{g}}} \end{bmatrix}\end{aligned} \tag{3.3}$$

The quantity $\boldsymbol{r}'\boldsymbol{r}$ is a robust estimate of the variance of the estimate $\hat{\beta}$ for the stratified log odds ratio (Wei, Lin and Weissfeld, 1989). The quantity $\boldsymbol{V}_{\hat{\beta},\bar{\boldsymbol{g}}} = \sum_{h=1}^{q}(\boldsymbol{C}'_{h1}\boldsymbol{r}_{h1} - \boldsymbol{C}'_{h2}\boldsymbol{r}_{h2})$ provides the estimated covariance matrix for $\hat{\beta}$ and $(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)$, and the quantity $\boldsymbol{V}_{\bar{\boldsymbol{g}}} = \sum_{h=1}^{q}\sum_{i=1}^{2}\boldsymbol{C}'_{hi}\boldsymbol{C}_{hi}$ is an unbiased estimate for the covariance matrix for $(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)$. The covariance structure for the covariates presented here is similar to the structure of $\boldsymbol{V}_w$ in Appendix II of Koch *et. al.* with appropriate weighting across the strata, and it is an unbiased estimate for the covariance structure under the assumption that patients are like a simple random sample from a stratified population (Koch et al., 1998). Two other specifications for the covariance matrix for $\boldsymbol{f}$, $\tilde{\boldsymbol{V}}_{\boldsymbol{f}}$ and $\bar{\boldsymbol{V}}_{\boldsymbol{f}}$, are presented in Appendices (A.1.1) and (A.2), respectively.

A covariate-adjusted estimate of β can then be obtained using weighted least squares methodology by forcing the difference in means for the P covariables to zero. The model $\boldsymbol{Z} = (1\ \boldsymbol{0}'_P)'$ can be fit to $\boldsymbol{f}$, where $\boldsymbol{0}_P$ is a zero vector of length P. The estimate of the

log odds ratio which is adjusted for covariates and for strata is b given in (3.4)

$$\begin{aligned} b &= (\boldsymbol{Z}'\boldsymbol{V}_{\boldsymbol{f}}^{-\mathbf{1}}\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{V}_{\boldsymbol{f}}^{-\mathbf{1}}\boldsymbol{f} \\ &= \hat{\beta} - \boldsymbol{V}'_{\hat{\beta},\bar{\boldsymbol{g}}}\boldsymbol{V}_{\bar{\boldsymbol{g}}}^{-\mathbf{1}}\bar{\boldsymbol{g}} \end{aligned} \tag{3.4}$$

and a consistent estimator for the variance of b, V_b is in (3.5).

$$\begin{aligned} V_b &= (\boldsymbol{Z}'\boldsymbol{V}_{\boldsymbol{f}}^{-\mathbf{1}}\boldsymbol{Z})^{-1} \\ &= v_{\hat{\beta}} - \boldsymbol{V}'_{\hat{\beta},\bar{\boldsymbol{g}}}\boldsymbol{V}_{\bar{\boldsymbol{g}}}^{-\mathbf{1}}\boldsymbol{V}_{\hat{\beta},\bar{\boldsymbol{g}}} \end{aligned} \tag{3.5}$$

The estimator b has an approximate normal distribution when each treatment group has at least minimal sample sizes within each stratum (e.g. $n_{hi} \geq 5$) and at least moderate overall sample size (e.g. ≥ 50) so that $\boldsymbol{f}$ has an approximately multivariate normal distribution. This methodology relies upon the assumption of a valid stratified randomization, under which no differences between the stratification adjusted means of the covariables are expected between the two treatment groups. Random imbalances in these means between the two groups can be assessed with the statistic Q_0 in (3.6).

$$Q_0 = (\boldsymbol{f} - \boldsymbol{Z}b)'\boldsymbol{V}_{\boldsymbol{f}}^{-\mathbf{1}}(\boldsymbol{f} - \boldsymbol{Z}b) \tag{3.6}$$

This statistic approximately has the chi-squared distribution with P degrees of freedom.

3.2.2 Hybrid Methodology for Time-to-Event Outcomes with Stratification

Let T_{hj} be the observed event time (or censoring time if no event) for the jth individual ($j = 1, \ldots, n_h$) from the hth stratum ($h = 1, \ldots, q$). The stratified treatment-only

Cox proportional hazards model for the hazard of an event at time t is given in (3.7).

$$\lambda(t; x_{hj}) = \lambda_{oh}(t) exp(\beta x_{hj}) \tag{3.7}$$

Here, $\lambda_{oh}(t)$ represents the underlying baseline hazard function for stratum h, and x_{hj} is an indicator of the test treatment (vs. comparator). As in the case for dichotomous outcomes, the vector of DFBETA residuals $\boldsymbol{r} = (r_1, \ldots, r_n)'$ can be obtained from the model fitting on the n individuals. The DFBETA r_j represents the absolute change in the estimated log hazard ratio $\hat{\beta}$ when the jth individual is removed from the analysis. Through the methods of Wei *et. al.*, a robust estimate of the variance of $\hat{\beta}$ is provided by the quantity $\boldsymbol{r'r}$ (Wei, Lin and Weissfeld, 1989). Thus, if $\bar{\boldsymbol{g}} = (\bar{\boldsymbol{x}}_{\boldsymbol{1}} - \bar{\boldsymbol{x}}_{\boldsymbol{2}})$ is defined to be the weighted difference in means of the covariables, the quantities $\boldsymbol{V}_{\bar{\boldsymbol{g}}}$ and $\boldsymbol{V}'_{\hat{\beta},\bar{\boldsymbol{g}}}$ are used to estimate a covariance matrix $\boldsymbol{V}_{\boldsymbol{f}}$ for the vector $\boldsymbol{f} = (\hat{\beta}, \bar{\boldsymbol{g}}')$ as in (3.3).

Weighted least squares then provides a covariate-adjusted estimate of β by forcing the differences in the stratified means of the P covariables to zero using the model $\boldsymbol{Z} = (1 \ \boldsymbol{0}'_P)'$. This estimate, which is also adjusted for the stratification factor, is given as $b = (\boldsymbol{Z'V_f^{-1}Z})^{-1}\boldsymbol{Z'V_f^{-1}f}$, and its estimated covariance matrix is given as $V_b = (\boldsymbol{Z'V_f^{-1}Z})^{-1}$. As previously noted for its counterpart in (3.4), the estimate b approximately has a normal distribution when the sample sizes in each group are large enough so that $\boldsymbol{f}$ has an approximately multivariate normal distribution. Also, random imbalances in the covariable means between the two groups can be assessed with the statistic $Q_0 = (\boldsymbol{f} - \boldsymbol{Z}b)'\boldsymbol{V_f^{-1}}(\boldsymbol{f} - \boldsymbol{Z}b)$, and this statistic approximately has the chi-squared distribution with P degrees of freedom.

3.3 Examples

3.3.1 Dichotomous Outcome: Neurologic Disorder Data

Researchers observed 431 events on 959 patients randomized to one of four treatment groups (placebo, 50 mg, 100 mg, 200 mg) for management of unfavorable outcomes for an incurable neurologic disorder. The primary event of interest was defined as time until progression of disease. Results of the main analysis of the data have been reported elsewhere (Lacomblez et al., 1996). Tangen and Koch (1999) provide discussion of implenting nonparametric randomization-based covariance adjustment to these data using Wilcoxon or logrank scores. For implementation of the NPANCOVA methods as they relate to dichotomous outcomes in this chapter, the primary outcome for this analysis was the Yes/No outcome of whether a patient progressed by 12 months after randomization. Tangen and Koch (2001) have discussion of dose-response relationships using nonparametric randomization-based covariance adjustment for this dichotomized outcome. Six patients who were censored prior to twelve months were managed as progression-free survivors. The 100 mg and 200 mg treatment groups were pooled (n=480) and compared to the placebo group (n=242). The 50 mg treatment group was omitted from the analysis. Strata were created as the cross-classification of geographic region and site of disease onset so as to create six strata.

A subset of six baseline covariates were chosen from a set of 24 candidate baseline covariates on the basis of a stepwise conditional logistic regression model (stratifying on geographic region, site of disease onset, and pooled treatment group: 100 mg + 200 mg vs. placebo) with an entry criterion of $p < 0.005$. This method produced a set of covariates predictive of having an event by 12 months in a manner independent of treatment assignment. Selected baseline covariates included age, disease duration (years), weight, and three additional continuous measures of neurologic and musculoskeletal function.

The overall missingness percentage for these covariates was 1.55%, and mean values for the covariates (over all 959 original patients) were imputed for these missing values.

Table 3.1: Progression by 12 Months for Patients with a Neurologic Disorder from Example 3.3.1 who Received Test Treatment (100 mg or 200 mg) or Placebo

		Disease Progression by 12 Months		
Strata	Treatment	Yes (Row %)	No	Total
1	Test	48 (31.2)	106	154
	Placebo	30 (37.5)	50	80
2	Test	24 (36.9)	41	65
	Placebo	20 (57.1)	15	35
3	Test	18 (16.1)	94	112
	Placebo	18 (32.1)	38	56
4	Test	19 (35.8)	34	53
	Placebo	10 (40.0)	15	25
5	Test	16 (25.0)	48	64
	Placebo	8 (25.0)	24	32
6	Test	4 (12.5)	28	32
	Placebo	4 (28.6)	10	14
		219	503	722

The counts and proportions of patients (by treatment and stratum) with disease progression by 12 months is presented in Table 3.1. Strata 1, 4, and 5 appear to have somewhat similar proportions of events between test treatment and placebo, but there is a lower proportion of events for the test treatment (vs. placebo) in strata 2, 3, and 6.

Results of the analysis are presented in Table 3.2. The stratified analysis without adjusting for covariates produced an odds ratio estimate of 0.6206, indicating the odds of having disease progression by 12 months were about 38% smaller for those on test treatment vs. placebo. This estimate had a corresponding p-value of 0.0057, indicating statistical significance at the 0.05 level. After adjustment for the baseline covariates, the NPANCOVA analysis showed a standard error reduction of 14.8% for the log odds ratio over the stratified analysis (from 0.1725 to 0.1470). The NPANCOVA adjusted

odds ratio estimate (0.5774) moved somewhat further away from the null than the stratified estimate of 0.6206, but the two values were reasonably similar and represent an odds ratio that is like a stratified population-averaged estimate. The NPANCOVA odds ratio estimate was in between the stratified estimate and the conditional logistic regression estimate (0.4579), but the larger standard error for the conditional logistic regression estimate (0.2091) resulted in the two analyses having the same p-value (0.0002). The conditional logistic regression estimate of 0.4579 was further away from the null than the stratified or NPANCOVA estimates, and this is likely due to it being a conditional estimate which is interpretable through the comparison of individuals who share the same values of both the strata and the covariates. The assessment for random imbalances in the covariate distributions had a p-value of 0.4106, and this p-value did not contradict the similarity of the covariate distributions between the treatment groups.

3.3.2 Time-to-Event Outcome: Neurologic Disorder Data

The data from Example 3.3.1 are reconsidered with application of NPANCOVA methods to the time-to-event outcome of progression-free survival over the main follow-up period of 18 months. The same six strata were used in the analysis, and the sample sizes and number of events for the strata are presented in Table 3.3. For the Cox regression, the Efron method for ties was used (Efron, 1977).

A subset of six baseline covariates were chosen from a set of 24 candidate baseline covariates on the basis of a stepwise Cox proportional hazards model (stratifying on geographic region, site of disease onset, and treatment) with an entry criterion of $p < 0.005$. This produced a set of covariates predictive of time to progression in a manner independent of treatment assignment. Selected baseline covariates included age, disease duration (years), and the same three additional continuous measures of neurologic and

musculoskeletal function which were included for the analysis for the dichotomized outcome in Example 3.3.1. While the baseline weight did not meet the criteria for entry for the stepwise Cox model (p=0.0081 > 0.005), this variable was considered predictive of the outcome and was included so that the same six covariates were considered for both the dichotomized outcome and the time-to-event outcome.

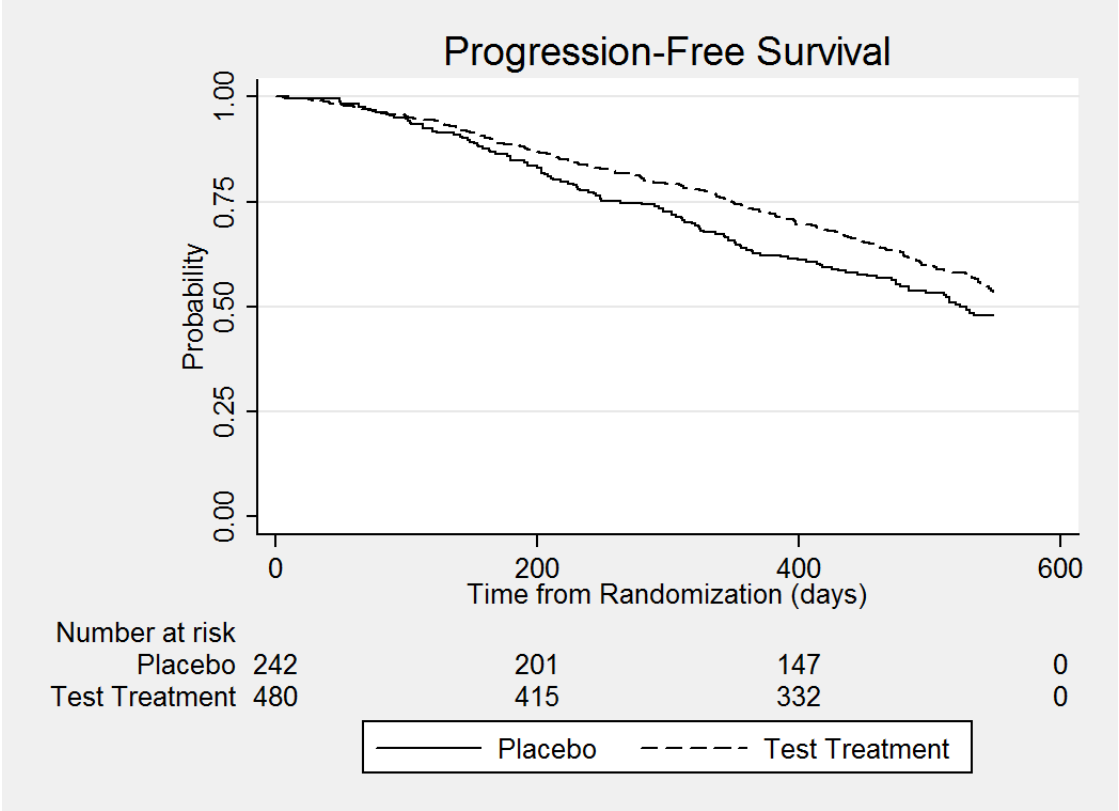

Figure 3.1: Kaplan-Meier Estimates of Progression-Free Survival for Neurologic Data from Example 3.3.2

Kaplan-Meier curves are displayed in Figure 3.1 for describing time to progression for the pooled treatment (100 mg + 200 mg) vs. placebo. This unadjusted analysis (with strata and covariables ignored) shows a separation of the survival curves, and this is reflected in the p-value for the logrank test (p=0.0480) and Wilcoxon test (p=0.0263). When considering the six strata representing geographic region/site of disease onset, the stratified logrank test had p=0.0322.

Table 3.4 displays results from stratified analyses and stratified analyses adjusted for the six covariates. The stratified analysis using the Cox model-based variance is provided for reference. The stratified analysis using the robust variance of Wei *et. al.* produced a hazard ratio of 0.7808, while the NPANCOVA and Cox adjusted analyses produced hazard ratios further away from the null (0.7392 and 0.6170, respectively). The hazard ratio for the NPANCOVA adjusted analysis was reasonably similar to the stratified analysis, as it is like a stratified population-averaged estimate. The hazard ratio for the Cox adjusted analysis is more like a conditional estimate which compares individuals from the different treatment groups who share the same values of the covariates and strata.

The p-value of 0.0316 from the stratified analysis was similar to the p-value from the stratified logrank test (p=0.0322), indicating that the analysis which assumes proportional hazards on the treatment variable produced similar results as the nonparametric test. Adjustment for covariates via NPANCOVA resulted in a 18.2% reduction in the standard error for the log hazard ratio, from 0.1151 (stratified) to 0.0942 (NPANCOVA). The standard error from the Cox model was larger than that of the stratified analysis, but the stronger estimate of the log hazard ratio resulted in the smallest p-value of < 0.0001. While the p-values for the stratified analysis and the NPANCOVA analysis were both below a significance level of 0.05, the covariance adjustment provided a reduction in the standard error for the log hazard ratio and stronger evidence of efficacy (through p=0.0013). The assessment of random imbalances in the covariates had $Q_0 = 6.1131$ with six degrees of freedom, and p=0.4106. This p-value did not contradict the similarity of distributions in the covariates between the treatment groups.

3.3.3 Dichotomous Outcome: Osteoarthritis Data

This example uses data from a randomized crossover trial for osteoarthritis of the hip or knee (Pincus, Koch and Sokka, 2001). Researchers randomized 227 patients to receive either experimental treatment (A) in the first period followed by active control (C) in the second period (or C followed by A), with washout in between periods. A screening visit was conducted (Visit 1), followed by a 3-7 day washout period before the randomization at the baseline visit (Visit 2). After six weeks of treatment (the end of the first period, Visit 3), patients returned to have pain assessments. Patients were given the option to continue to the second treatment period. If they consented, a 3-7 day washout period occurred, followed by the second baseline visit (Visit 4) where they received a six-week supply of the treatment they did not receive in the first period. Pain assessments were conducted at the end of this six-week treatment period (Visit 5). A questionnaire used to assess pain was the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC). This questionnaire has 24 items with each item being scored using the following Likert scale: 0=None, 1=Mild, 2=Moderate, 3=Severe, 4=Extreme. The WOMAC index is calculated as (sum of item scores)/96 (to produce an index range of 0-100) with higher values indicating worse pain.

For this example, only data from the screening visit (Visit 1), the baseline visit (Visit 2), and the end of the first period (Visit 3) were considered. There were 218 patients who had available data from all three visits. Study site was used to form the strata. The 218 patients included for analysis were enrolled at twelve study sites, and the sample sizes for the strata are listed in Table 3.5. The continuous WOMAC outcome was dichotomized as a variable which took the value 1 if the score was ≤ 25 at the end of the first period (Visit 3) or 0 if the score was > 25 at the end of the first period. This cut point was chosen to represent an index for an individual answering 'Mild' for all 24 items on the questionnaire. Adjustment variables for analysis of covariance included sex,

continuous baseline age, continuous screening outcome score, and continuous baseline outcome score. For the analyses, ten strata were used after pooling the 7, 6, and 3 patients from strata 1, 5, and 7, respectively.

Table 3.6 displays the results of the stratified analyses for the WOMAC dichotomized outcome. When considering the WOMAC scale, the stratified odds ratio comparing the odds of a WOMAC score ≤ 25 for the experimental treatment group to the odds for the active control was 2.0026. After adjustment via NPANCOVA, the odds ratio moved closer to the null with a value of 1.7044. Adjustment via conditional logistic regression produced an estimated odds ratio of 2.1135, which was larger than the stratified odds ratio and further from the null. Finally, adjustment for covariables and strata as fixed effects in an unconditional logistic regression produced an odds ratio of 2.2122, which was the furthest from the null and may be reflecting bias due to some sites having very small sample sizes. The odds ratio estimate for NPANCOVA is more like a stratified population-averaged estimate of the odds ratio, whereas the estimates from conditional or unconditional logistic regression reflect conditional estimates of the odds ratio for comparing treatment groups for individuals from the same strata who share the same values of the other covariates.

The standard error for the stratified log odds ratio for the WOMAC outcome was 0.2830. With adjustment via NPANCOVA, there was a 17.5% reduction in the standard error to 0.2335. For the conditional and unconditional logistic regression models, the standard errors increased over their counterparts without covariables, with values of 0.3436 and 0.3535, respectively. While the standard errors for these methods were larger than the stratified approach, the estimated log odds ratios were larger and resulted in p-values of 0.0276 and 0.0231. While the stratified and NPANCOVA p-values were smaller at 0.0141 and 0.0224, respectively, all the p-values for this example were statistically significant at the 0.05 level and indicated efficacy of the experimental treatment. The

assessment of random imbalance of the covariates had $Q_0 = 1.1792$ with four degrees of freedom and had p=0.8815. This p-value did not contradict the distributions of the covariates being similar between the two treatment groups.

3.4 Discussion

The hybrid methdology presented in this chapter is useful in a regulatory setting where a conditional analysis with covariate adjustment is desired when analyzing a dichotomous outcome, but the assumptions of a parametric regression model can not be verified prior to data collection. While the procedure does make use of a parametric treatment-only model, the score test for the hypothesis $H_0 : \beta = 0$ in the treatment-only conditional logistic regression model is essentially identical to the Mantel-Haenszel test which only relies on a valid randomization. Thus, the hybrid methodology does not require additional modeling assumptions of conditional logistic regression such as correct functional form for the covariates. The covariate-adjusted estimate of the treatment effect which is produced by the hybrid methodology is a stratified population-averaged estimate, in contrast with the conditional logistic regression estimate which is interpretable as a treatment effect for individuals who share the same values of the strata and the covariates. It is recommended that the hybrid methodology be used to obtain an adjusted treatment effect for the primary analysis. Subsequent assessment of interactions or subgroup differences could be performed via supportive conditional logistic regression analyses.

In the case of time-to-event data, the stratified NPANCOVA analysis could be used when covariate adjustment is desired but the assumptions of a stratified Cox proportional hazards model may not hold. Specifically, the appropriate functional form of the covariates can not be verified *a priori*, and most importantly, the proportional hazards assumption can not be shown to hold for the covariates before data is collected.

The methodology does use a treatment-only Cox model, but when using the Efron method for handling ties, the p-value from the treatment-only model is similar to that from the nonparametric stratified logrank test. The NPANCOVA analysis does provide the additional robustness of not requiring the proportional hazards assumption for the covariates. Example 3.3.2 reflects this, as further analysis using scaled Schoenfeld residuals (not shown) appeared to show violation of proportional hazards for one of the baseline covariates.

The covariance matrix $\boldsymbol{V_f}$ for $\boldsymbol{f}$ presented in Section 3.2.1 is an extension of the covariance matrix in Koch *et. al.* (1998) for the stratified setting. It assumes that the patients are like a random sample from a corresponding stratified population, and it is formed in a way that is similar to the covariance matrix in Saville and Koch (Saville and Koch, 2013). Another specification for the covariance matrix of $\boldsymbol{f}$, $\tilde{\boldsymbol{V}}_{\boldsymbol{f}}$, is developed using multivariate U-statistics and is presented in Appendix A.1.1. Results of the analyses in Appendix A.3 indicate that the covariate-adjusted NPANCOVA estimates and standard errors are similar, despite some differences in the numerical values of the covariance matrices. In particular, the comparisons of the components of $\tilde{\boldsymbol{V}}_{\boldsymbol{f}}$ with $\boldsymbol{V_f}$ in Appendix A.1.2 suggest more similarity in the two covariance matrices if stratum sizes are larger (as in the Neurologic Disorder example) than if they are around 10-20 (as in the Osteoarthritis example). However, whether $\boldsymbol{V_f}$ or $\tilde{\boldsymbol{V}}_{\boldsymbol{f}}$ should be used may be irrelevant; the estimates, standard errors, and p-values were very similar, and the inference for the treatment effect was the same. A third specification, $\bar{\boldsymbol{V}}_{\boldsymbol{f}}$, contains components including the model-based variance and the covariance structure for the weighted difference in means of the covariates $\boldsymbol{V}_{\bar{\boldsymbol{g}}}$ which pertains to $\boldsymbol{V_f}$. The use of $\bar{\boldsymbol{V}}_{\boldsymbol{f}}$ resulted in estimates and standard errors that were very similar to those obtained from using $\boldsymbol{V_f}$ or $\tilde{\boldsymbol{V}}_{\boldsymbol{f}}$, and, for the dichotomized neurologic disorder outcome, the use of $\bar{\boldsymbol{V}}_{\boldsymbol{f}}$ resulted in a standard error that was smaller than when using the other two

specifications.

The DFBETA residuals obtained for the analyses are part of standard output in most software packages. The residuals for the examples in this chapter were obtained from PROC PHREG in SAS. While the main intent of this procedure is for fitting proportional hazards models for time-to-event data, it was used to fit the conditional logistic regression models (via the 'ties=discrete' option in SAS). This approach is reasonable since the partial likelihood for time-to-event data reduces to the conditional likelihood for a dichotomous outcome when all events are assumed to occur at the same time. The DFBETA residuals output by PROC PHREG are the product of the estimated variance of $\hat{\beta}$ and the score residuals for the ith subject, and these residuals are produced for all observations with complete data for the time-to-event or dichotomous outcome. The sum of squares of the DFBETA residuals gives a robust variance estimate of the treatment effect $\hat{\beta}$ (Wei, Lin and Weissfeld, 1989). Storer and Crowley developed DFBETA residuals for conditional likelihoods, and these can be requested through PROC LOGISTIC in SAS (Storer and Crowley, 1985). These residuals are based on the one-step method by starting at the MLE $\hat{\beta}$, deleting the jth individual from the hth stratum, and making the next Newton-Raphson step to obtain $\hat{\beta}_{(hj)}$ (Storer and Crowley, 1985) (Pregibon, 1981). Then, the DFBETA residual is given as $(\hat{\beta} - \hat{\beta}_{(hj)})/\hat{\sigma}$ where $\hat{\sigma}$ is the estimated standard error of $\hat{\beta}$ from the full data set. Unlike the DFBETA residuals produced from PROC PHREG, the residuals from PROC LOGISTIC are only produced for observations from informative strata, and the robustness of the sum of squares of these DFBETA residuals for variance estimation is less clear.

For dichotomous outcomes, the methods in Chapter 2 are intended for very small strata, and the methods in this chapter are intended for larger strata of size ≥ 10. For time-to-event outcomes, the methods here are useful for a reasonably small number

of strata as would be appropriate in a stratified time-to-event analysis to ensure the number of events is large enough for each stratum. In the next chapter, the hybrid methodology of Saville and Koch is extended to the case of more than two randomized treatment groups for time-to-event outcomes.

Table 3.2: Stratified Analysis for Dichotomized Outcome from Example 3.3.1 for Neurologic Disorder Data

Method	Log OR	S.E.	OR	95% CI		p-value
				Lower	Upper	
Stratified, Robust Variance[a]	-0.4771	0.1725	0.6206	0.4426	0.8702	0.0057
NPANCOVA Adjusted[b]	-0.5492	0.1470	0.5774	0.4329	0.7702	0.0002
Stratified, Model-Based Variance	-0.4771	0.1704	0.6206	0.4444	0.8667	0.0049[c]
Conditional Logistic Regression Adjusted[b]	-0.7810	0.2091	0.4579	0.3040	0.6899	0.0002[c]
Unconditional Logistic Regression Adjusted[b]	-0.7895	0.2103	0.4541	0.3007	0.6857	0.0001[c]

[a]Robust variance based on Wei *et. al.*

[b]Adjusted for six continuous baseline covariates

[c]P-value from score test

Table 3.3: Strata Sizes and Number of Events for the 722 Patients Enrolled on the Placebo, 100 mg, or 200 mg Arms of the Neurologic Disorder Trial for Example 3.3.2

Strata	Sample Size (%)	Number of Events (%)
1	234 (32.3)	113 (34.8)
2	100 (13.9)	69 (21.2)
3	168 (23.3)	52 (16.0)
4	78 (10.8)	42 (12.9)
5	96 (13.3)	29 (8.9)
6	46 (6.4)	20 (6.2)
	722 (100.0)	325 (100.0)

Table 3.4: Stratified Analysis for Time-to-Event Outcome from Example 3.3.2 for Neurologic Disorder Data

Method	Log HR	S.E.	HR	95% CI		p-value
				Lower	Upper	
Stratified, Robust Variance[a]	-0.2475	0.1151	0.7808	0.6231	0.9785	0.0316
NPANCOVA Adjusted[b]	-0.3021	0.0942	0.7392	0.6146	0.8891	0.0013
Stratified, Model-Based Variance	-0.2475	0.1154	0.7808	0.6227	0.9790	0.0316[c]
Cox Adjusted[b]	-0.4829	0.1193	0.6170	0.4884	0.7795	<0.0001

[a]Robust variance based on Wei, *et. al.*
[b]Adjusted for six continuous baseline covariates
[c]P-value from score test

Table 3.5: Strata Sizes for the 218 Patients with Data for Visits 1, 2, and 3 in the Osteoarthritis Randomized Crossover Trial for Example 3.3.3

Study Site	Sample Size (%)
1	7 (3.2)
2	21 (9.6)
3	10 (4.6)
4	30 (13.8)
5	6 (2.8)
6	15 (6.9)
7	3 (1.4)
8	22 (10.1)
9	19 (8.7)
10	45 (20.6)
11	21 (9.6)
12	19 (8.7)
	218 (100.0)

Table 3.6: Results from Example 3.3.3 for Osteoarthritis Data for the WOMAC Dichotomized Outcome

Method	Log OR	S.E.	OR	95% CI		p-value
				Lower	Upper	
Stratified, Robust Variance[a]	0.6945	0.2830	2.0026	1.1501	3.4872	0.0141
NPANCOVA Adjusted[b]	0.5332	0.2335	1.7044	1.0785	2.6936	0.0224
Stratified, Model-Based Variance	0.6945	0.2762	2.0026	1.1655	3.4413	0.0114[d]
Conditional Logistic Regression Adjusted[b]	0.7484	0.3436	2.1135	1.0777	4.1448	0.0276[d]
Unconditional Logistic Regression Adjusted[b,c]	0.7940	0.3535	2.2122	1.1064	4.4232	0.0231[d]

[a] Robust variance based on Wei *et. al.*

[b] Adjusted for sex, Visit 2 continuous age, Visit 1 continuous outcome, Visit 2 continuous outcome

[c] Adjustment for strata through 11 indicator variables

[d] P-values from score test

Chapter 4

Covariate-adjusted Log Hazard Ratios using Cox Proportional Hazards Regression and Nonparametric Randomization-Based ANCOVA for More Than Two Treatments

4.1 Introduction

Analysis of covariance for time-to-event outcomes may occur through the Cox proportional hazards model (Cox, 1972). This model allows for estimation of a log hazard ratio (or ratios) for comparing two (or more) treatments while adjusting for covariables. The main assumption of using the Cox model is that the covariates (including treatment) satisfy proportional hazards. That is, the multiplicative effect of a treatment (or influence of a covariate) on the underlying hazard function is assumed constant over time. Another assumption is that the covariates themselves have the correct functional form. For continuous covariates, this involves understanding of the relationship between the covariate and the log hazard function. In the regulatory setting, analyses must be specified in a protocol prior to data collection. Thus, the statistical assumptions of a

Cox model are not verifiable *a priori* and its use for the primary analysis of a randomized clinical trial in this setting may lead to questionable results if the assumptions do not hold.

The general NPANCOVA strategy has been applied to continuous, dichotomous, and ordinal outcomes (Koch et al., 1998), (Tangen and Koch, 1999*a*), (Kawaguchi, Koch and Wang, 2011). Tangen and Koch (Tangen and Koch, 1999*b*) applied the methodology to time-to-event outcomes via differences in mean logrank or Wilcoxon scores, but this does not provide a log hazard ratio. A subsequent method by Moodie *et. al.* (Moodie et al., 2011) provided a log hazard ratio based on the means of indicators for risk and survival for pre-determined time intervals, but this does not have the optimal properties of the log hazard ratio provided from the Cox model.

Methods introduced by Saville and Koch (Saville and Koch, 2013) use the estimated log hazard ratios from a treatment-only Cox proportional hazards model for multivariate time-to-event outcomes as part of covariance adjustment provided through NPANCOVA. The compound difference vector contains the unadjusted log hazard ratios from the treatment-only Cox model as well as the difference in means of the covariables between the two treatment groups. The covariance matrix for the compound difference is obtained through three components: (1) A robust estimator for the covariance matrix of the estimated log hazard ratios provided through the DFBETA residuals from the treatment-only Cox model (Wei, Lin and Weissfeld, 1989), (2) the estimated covariance matrix of the difference in means of covariables derived through multivariate U-statistics, and (3) the covariance matrix of the estimated log hazard ratios and the difference in means of the covariables. The compound vector and its covariance matrix are then used in weighted least squares to obtain a covariate-adjusted estimate of the log hazard ratios.

The NPANCOVA strategies described thus far have been created for randomization

to two treatment groups. Since many clinical trials have randomization to more than two groups either because of multiple treatments or multiple doses, there is interest in adapting the methodology to these settings. This chapter introduces methodology for extension of the treatment-only Cox proportional hazards model and NPANCOVA (Saville and Koch, 2013) to more than two treatment groups. Covariate-adjusted log hazard ratios are estimated for pairwise comparisons of each group with a referent group. Where the multiple treatment groups correspond to multiple doses of the same treatment, global assessment of an experimental treatment versus control may proceed through testing a linear contrast (i.e. trend) for better response with increasing dose. Subsequent assessment of pairwise differences in treatment groups could then occur with appropriate management of multiplicity (Tangen and Koch, 2001). An additional extension is provided for the estimation of treatment effects which have departures from proportional hazards in the sense of time by treatment interaction for their variation across pre-determined time intervals. This methodology has utility for providing covariate-adjusted estimates of log hazard ratios for pairwise comparisons of treatments and assessing the extent of their homogeneity. The methods are presented in Section 4.2, followed by an example and subsequent discussion.

4.2 Methods

4.2.1 Methodology for More than Two Treatment Groups and One Time-to-Event Outcome

Let T_j be the failure time (or censoring time if no event) for patient j and $z_{ij} = 1$ if patient j is randomized to treatment group i ($i = 1, \ldots, s-1$) and $z_{ij} = 0$ otherwise. This construction creates $s-1$ indicator variables for treatment groups $i = 1, \ldots, s-1$ with the sth treatment group being the referent group. The unadjusted

Cox proportional hazards model for the time-to-event outcome is in (4.1).

$$\lambda_j(t; z_{1j}, \ldots, z_{(s-1)j}) = \lambda_0(t)\exp(\beta_1 z_{1j} + \ldots + \beta_{s-1} z_{(s-1)j}) \tag{4.1}$$

Here, $\lambda_j(t; z_{1j}, \ldots, z_{(s-1)j})$ is the hazard function for patient j and $\lambda_0(t)$ is the underlying baseline hazard function for the sth treatment group.

With $n = \sum_{i=1}^{s} n_i$ individuals, let $\boldsymbol{r}^{(i)} = (r_1^{(i)}, \ldots, r_n^{(i)})'$ be the vector of n unstandardized DFBETA residuals from fitting the model in (4.1) for the comparison of the ith treatment group with the sth treatment group ($i = 1, \ldots, s-1$). The DFBETA residual $r_j^{(i)}$ is the approximate change in the log hazard ratio comparing the ith group to the sth group when the jth individual is omitted. Wei *et. al.* (1989) showed $\boldsymbol{r}^{(i)'}\boldsymbol{r}^{(i)}$ provides a robust variance estimate of the estimated log hazard ratio comparing the ith group to the sth group (Wei, Lin and Weissfeld, 1989). With $\boldsymbol{R} = (\boldsymbol{r}^{(1)}, \ldots, \boldsymbol{r}^{(s-1)})$ representing the $n \times (s-1)$ matrix of the DFBETA residuals, the matrix $\boldsymbol{V}_{\hat{\boldsymbol{\beta}}} = \boldsymbol{R}'\boldsymbol{R}$ is a robust covariance matrix estimate for the estimated log hazard ratios $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \ldots, \hat{\beta}_{s-1})'$ for the $s-1$ comparisons with the sth group.

Let $\boldsymbol{X}_i = (\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{iP})$ be the $n_i \times P$ matrix of P baseline covariates for the n_i individuals from the ith treatment group. With $\bar{\boldsymbol{x}}_i = (\bar{x}_{i1}, \ldots, \bar{x}_{iP})'$ representing the vector of covariate means for group i, the covariance matrix $\boldsymbol{V}_{\bar{\boldsymbol{x}}_i} = (\boldsymbol{X}_i - \mathbf{1}\bar{\boldsymbol{x}}_i')'(\boldsymbol{X}_i - \mathbf{1}\bar{\boldsymbol{x}}_i')/[n_i(n_i-1)] = \boldsymbol{A'}_i\boldsymbol{A}_i$ is the estimated covariance matrix of the covariate means for group i (with $\mathbf{1}$ being an $n_i \times 1$ vector of ones). Here, $\boldsymbol{A}_i = (\boldsymbol{X}_i - \mathbf{1}\bar{\boldsymbol{x}}_i')/\sqrt{n_i(n_i-1)}$.

Let $\boldsymbol{d} = (\hat{\boldsymbol{\beta}}', (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_s)', \ldots, (\bar{\boldsymbol{x}}_{s-1} - \bar{\boldsymbol{x}}_s)')' = (\hat{\boldsymbol{\beta}}', \boldsymbol{u}')'$ be the $((P+1)(s-1)) \times 1$ compound vector of estimated log hazard ratios and differences in baseline covariate means for the $s-1$ treatment comparisons. Then $\boldsymbol{V_d}$ in (4.2) is the $((P+1)(s-1)) \times ((P+1)(s-1))$ covariance matrix of $\boldsymbol{d}$ where $\boldsymbol{V}_{\hat{\boldsymbol{\beta}}} = \boldsymbol{R}'\boldsymbol{R}$ is the estimated covariance matrix for the estimated log hazard ratios, $\boldsymbol{V}_{\bar{\boldsymbol{x}}_i} + \boldsymbol{V}_{\bar{\boldsymbol{x}}_s}$ is the estimated covariance matrix for $(\bar{\boldsymbol{x}}_i - \bar{\boldsymbol{x}}_s)$, $\boldsymbol{V}_{\bar{\boldsymbol{x}}_s}$ is the estimated covariance matrix for $(\bar{\boldsymbol{x}}_i - \bar{\boldsymbol{x}}_s)$ with $(\bar{\boldsymbol{x}}_{i'} - \bar{\boldsymbol{x}}_s)$ where

$i \neq i'$, and $\boldsymbol{V}_{\hat{\boldsymbol{\beta}},\bar{\boldsymbol{x}}_{\boldsymbol{i}}} - \boldsymbol{V}_{\hat{\boldsymbol{\beta}},\bar{\boldsymbol{x}}_{\boldsymbol{s}}}$ is the estimated covariance matrix of $\hat{\boldsymbol{\beta}}$ and $(\bar{\boldsymbol{x}}_i - \bar{\boldsymbol{x}}_s)$. With $\boldsymbol{R}_i$ representing the $n_i \times (s-1)$ matrix of DFBETA residuals for the n_i individuals in treatment group i, the quantity $\boldsymbol{V}_{\hat{\boldsymbol{\beta}},\bar{\boldsymbol{x}}_{\boldsymbol{i}}} - \boldsymbol{V}_{\hat{\boldsymbol{\beta}},\bar{\boldsymbol{x}}_{\boldsymbol{s}}}$ is estimated by $\boldsymbol{R'}_i\boldsymbol{A}_i - \boldsymbol{R'}_s\boldsymbol{A}_s$. Details are provided in Appendix A.3.3.

$$
\begin{aligned}
\boldsymbol{V_d} &= \begin{bmatrix}
\boldsymbol{V}_{\hat{\boldsymbol{\beta}}} & \boldsymbol{V}_{\hat{\boldsymbol{\beta}},\bar{\boldsymbol{x}}_{\boldsymbol{1}}} - \boldsymbol{V}_{\hat{\boldsymbol{\beta}},\bar{\boldsymbol{x}}_{\boldsymbol{s}}} & \boldsymbol{V}_{\hat{\boldsymbol{\beta}},\bar{\boldsymbol{x}}_{\boldsymbol{2}}} - \boldsymbol{V}_{\hat{\boldsymbol{\beta}},\bar{\boldsymbol{x}}_{\boldsymbol{s}}} & \cdots & \boldsymbol{V}_{\hat{\boldsymbol{\beta}},\bar{\boldsymbol{x}}_{\boldsymbol{s-1}}} - \boldsymbol{V}_{\hat{\boldsymbol{\beta}},\bar{\boldsymbol{x}}_{\boldsymbol{s}}} \\
 & \boldsymbol{V}_{\bar{\boldsymbol{x}}_{\boldsymbol{1}}} + \boldsymbol{V}_{\bar{\boldsymbol{x}}_{\boldsymbol{s}}} & \boldsymbol{V}_{\bar{\boldsymbol{x}}_{\boldsymbol{s}}} & \cdots & \boldsymbol{V}_{\bar{\boldsymbol{x}}_{\boldsymbol{s}}} \\
 & & \ddots & \ddots & \vdots \\
 & & & \ddots & \boldsymbol{V}_{\bar{\boldsymbol{x}}_{\boldsymbol{s}}} \\
 & & & & \boldsymbol{V}_{\bar{\boldsymbol{x}}_{\boldsymbol{s-1}}} + \boldsymbol{V}_{\bar{\boldsymbol{x}}_{\boldsymbol{s}}}
\end{bmatrix} \\
&= \begin{bmatrix}
\boldsymbol{V}_{\hat{\boldsymbol{\beta}}} & \boldsymbol{V}_{\hat{\boldsymbol{\beta}},\boldsymbol{u}} \\
\boldsymbol{V'}_{\hat{\boldsymbol{\beta}},\boldsymbol{u}} & \boldsymbol{V}_{\boldsymbol{u}}
\end{bmatrix}
\end{aligned}
\tag{4.2}
$$

Weighted least squares is then used to produce covariate-adjusted estimates $\boldsymbol{b}$ of the log hazard ratios by forcing the difference in means for the covariables to zero on the basis of this expected structure from randomization. This has invocation for $\boldsymbol{d}$ by fitting the model $\boldsymbol{Z} = [\boldsymbol{I}_{(s-1)} \quad \boldsymbol{0}_{(s-1)\times P(s-1)}]'$ to $\boldsymbol{d}$, where $\boldsymbol{I}_{(s-1)}$ is a $(s-1)\times(s-1)$ identity matrix and $\boldsymbol{0}_{(s-1)\times P(s-1)}$ is a $(s-1)\times P(s-1)$ matrix of zeroes. The covariate-adjusted estimator $\boldsymbol{b}$ is given in (4.3).

$$
\begin{aligned}
\boldsymbol{b} &= (\boldsymbol{Z'}{\boldsymbol{V_d}}^{-1}\boldsymbol{Z})^{-1}\boldsymbol{Z'}{\boldsymbol{V_d}}^{-1}\boldsymbol{d} \\
&= \hat{\boldsymbol{\beta}} - \boldsymbol{V}_{\hat{\boldsymbol{\beta}},\boldsymbol{u}}\boldsymbol{V}_{\boldsymbol{u}}^{-\boldsymbol{1}}\boldsymbol{u}
\end{aligned}
\tag{4.3}
$$

and a consistent estimator for the covariance matrix of $\boldsymbol{b}$ is $\boldsymbol{V_b}$ in (4.4).

$$
\begin{aligned}
\boldsymbol{V_b} &= (\boldsymbol{Z'}{\boldsymbol{V_d}}^{-1}\boldsymbol{Z})^{-1} \\
&= \boldsymbol{V}_{\hat{\boldsymbol{\beta}}} - \boldsymbol{V}_{\hat{\boldsymbol{\beta}},\boldsymbol{u}}\boldsymbol{V}_{\boldsymbol{u}}^{-\boldsymbol{1}}\boldsymbol{V'}_{\hat{\boldsymbol{\beta}},\boldsymbol{u}}
\end{aligned}
\tag{4.4}
$$

These estimators $\boldsymbol{b}$ have an approximate multivariate normal distribution when the sample sizes for each treatment group are sufficiently large for $\boldsymbol{d}$ to have an approximate multivariate normal distribution provided $\hat{\boldsymbol{\beta}}$ also has an approximate multivariate normal distribution. Individual hypothesis tests for the log hazard ratios comparing each group i with group s can be conducted using the statistic $b_i^2/v_{b,i}$ where b_i is the estimated adjusted log hazard ratio and $v_{b,i}$ is the estimated variance of b_i. This statistic, through the approximate normality of b_i, has an approximate chi-squared distribution with one degree of freedom. A $100(1-\alpha)\%$ confidence interval for each log hazard ratio may be obtained through $b_i \pm z_{1-\alpha/2}\sqrt{v_{b,i}}$ where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$th percentile of a standard normal distribution. As the differences in means of the covariables between treatment groups are expected to be zero due to randomization, the extent of random imbalances between groups for the means of the covariates may be evaluated through Q_0 in (4.5).

$$Q_0 = (\boldsymbol{d} - \boldsymbol{Z}\boldsymbol{b})'\boldsymbol{V}_{\boldsymbol{d}}{}^{-1}(\boldsymbol{d} - \boldsymbol{Z}\boldsymbol{b}) \tag{4.5}$$

This criterion approximately has the chi-squared distribution with $P(s-1)$ degrees of freedom.

In the case of the treatment groups being different doses of an experimental treatment vs. a control, Tangen and Koch (2001) provide guidance on assessment of treatment effects in a manner which appropriately controls the experimentwise significance level. Global assessment of efficacy for the treatment vs. control can first proceed with a linear contrast based on prior understanding of the underlying dose-response relationship. For the NPANCOVA setting, this can be achieved by the appropriate specification for a contrast $\boldsymbol{c}$. For example, with high, medium, and low dose groups compared to placebo as a referent group, if the high and medium doses were thought to have similar effect but perform better than low dose, one could test the average of high and medium doses versus placebo at significance level α as an initial global assessment through the

hypothesis $H_0 : \boldsymbol{c}'\boldsymbol{b} = 0$. Here, $\boldsymbol{c}' = (1\ 1\ 0)$ and $\boldsymbol{b}' = (b_1\ b_2\ b_3)$ so that b_1 represents the covariate-adjusted log hazard ratio for high dose vs. placebo, b_2 represents medium vs. placebo, and b_3 represents low vs. placebo. This contrast can be tested using the statistic $Q_t = (\boldsymbol{cb})'(\boldsymbol{cV}_b\boldsymbol{c}')^{-1}(\boldsymbol{cb})$. Q_t then has an approximate chi-squared distribution with one degree of freedom. Once an overall treatment effect has been established through a significant p-value for the global assessment, individual pairwise comparisons between the treatment groups could be conducted. Kong *et. al.* (2005) offer some strategies for managing subsequent testing after the global assessment. In the example, if the test of the average of high and medium doses (versus placebo) is statistically significant, one could then test both high vs. placebo and medium vs. placebo at the full α. If both tests are statistically significant, one can then test low vs. placebo, and if significant, one could proceed to test other comparisons for the treatment groups in a manner like that in Kong *et. al.* (2005). For example, one could test the average of high and medium doses versus low; and if its result was significant, then one could test high versus low, medium versus low, and high versus medium at the full α on the basis of closed testing principles (Kong et al., 2005).

Alternatively, one could assess overall homogeneity of the log hazard ratios through the statistic $Q_h = (\boldsymbol{Lb})'(\boldsymbol{LV}_b\boldsymbol{L}')^{-1}(\boldsymbol{Lb})$ where $\boldsymbol{L} = \boldsymbol{I}_{s-1}$. Q_h has an approximate chi-squared distribution with $s-1$ degrees of freedom. Rejection of the null hypothesis $H_0 : \boldsymbol{Lb} = \boldsymbol{0}$ would then allow for subsequent assessment of the pairwise treatment comparisons with appropriate consideration for multiplicity.

4.2.2 Estimation of Log Hazard Ratios for Time-Varying Treatment Effects with More than Two Treatment Groups

In some cases of non-proportional hazards for the treatment variable, a Cox model may be fit which accommodates the treatment indicators as time-varying covariates.

When hazard ratios are thought to be different before and after a fixed time t^*, the time-dependence can be modeled using an indicator variable $w(t)$ where $w(t) = 1$ if $t > t^*$ and $w(t) = 0$ if $t \leq t^*$. The Cox model then takes the form in (4.6).

$$\lambda_j(t; z_{1j}, \ldots, z_{(s-1)j}, w(t)) = \lambda_0(t)\exp(\tilde{\beta}_1 z_{1j} + \ldots + \tilde{\beta}_{s-1} z_{(s-1)j} + \gamma_1 w(t) z_{1j} + \ldots + \gamma_{s-1} w(t) z_{(s-1)j}) \quad (4.6)$$

The vector of log hazard ratios for the $s-1$ treatment comparisons for time $[0, t^*]$ as well the $s-1$ comparisons for time $t > t^*$ can be represented by the vector $\hat{\tilde{\boldsymbol{\beta}}} = (\hat{\tilde{\beta}}_1, \ldots, \hat{\tilde{\beta}}_{s-1}, (\hat{\tilde{\beta}}_1 + \hat{\gamma}_1), (\hat{\tilde{\beta}}_2 + \hat{\gamma}_2), \ldots, (\hat{\tilde{\beta}}_{s-1} + \hat{\gamma}_{s-1}))'$. The vector $\tilde{\boldsymbol{d}}$ can be created as in Section 4.2.1 where $\hat{\tilde{\boldsymbol{\beta}}}$ replaces $\hat{\boldsymbol{\beta}}$.

The covariance matrix $\boldsymbol{V}_{\tilde{\boldsymbol{d}}}$ can be constructed in a similar manner as in Section 4.2.1, where $\boldsymbol{V}_{\hat{\tilde{\boldsymbol{\beta}}}}$ replaces $\boldsymbol{V}_{\hat{\boldsymbol{\beta}}}$ and $\boldsymbol{V}_{\hat{\tilde{\boldsymbol{\beta}}}, \bar{\boldsymbol{x}}_i}$ replaces $\boldsymbol{V}_{\hat{\boldsymbol{\beta}}, \bar{\boldsymbol{x}}_i}$. To obtain the $2(s-1) \times 2(s-1)$ covariance matrix $\boldsymbol{V}_{\hat{\tilde{\boldsymbol{\beta}}}}$, duplicate records are created in the data set for patients with $T_j > t^*$. Patients with $T_j \leq t^*$ have one record in the data set where the failure (or censoring) time is T_j. For patients with $T_j > t^*$, the first record denotes observation over $[0, t^*]$ with censoring at time t^* and the second record denotes observation over $(t^*, T_j]$ with failure (or censoring) at T_j. Let $\boldsymbol{R} = (\boldsymbol{Q}_1\ \boldsymbol{Q}_2)$ be the $n \times 2(s-1)$ matrix of DFBETA residuals where $\boldsymbol{Q}_1$ is the $n \times (s-1)$ matrix of DFBETA residuals from a treatment-only Cox proportional hazards model on records relating to $[0, t^*]$ and $\boldsymbol{Q}_2$ is the matrix of DFBETA residuals from a treatment-only Cox proportional hazards model on records relating to $(t^*, T_j]$ with zeroes replacing missing values in $\boldsymbol{Q}_2$ for patients with $T_j \leq t^*$ so as to create a $n \times (s-1)$ matrix. Then $\boldsymbol{V}_{\hat{\tilde{\boldsymbol{\beta}}}} = \boldsymbol{R}'\boldsymbol{R}$ is a robust covariance estimate of the log hazard ratios $\hat{\tilde{\boldsymbol{\beta}}}$. The $2(s-1) \times P$ matrices for the estimated covariances of $\hat{\tilde{\boldsymbol{\beta}}}$ and $(\bar{\boldsymbol{x}}_i - \bar{\boldsymbol{x}}_s)$ with $i = 1, \ldots, s-1$ are given by $\boldsymbol{V}_{\hat{\tilde{\boldsymbol{\beta}}}, \bar{\boldsymbol{x}}_i} - \boldsymbol{V}_{\hat{\tilde{\boldsymbol{\beta}}}, \bar{\boldsymbol{x}}_s} = \boldsymbol{R}'_i\boldsymbol{A}_i - \boldsymbol{R}'_s\boldsymbol{A}_s$ where $\boldsymbol{R}_i$ is the $n_i \times 2(s-1)$ matrix of DFBETA

residuals for the n_i individuals in treatment group i.

Weighted least squares may then be applied using the vector $\tilde{\boldsymbol{d}}$, the covariance matrix $\boldsymbol{V}_{\tilde{\boldsymbol{d}}}$, and the model $\tilde{\boldsymbol{Z}} = \boldsymbol{I}_{2(s-1)}$ to obtain the $2(s-1) \times 1$ vector $\tilde{\boldsymbol{b}}$ of covariate-adjusted estimates of the log hazard ratios for treatment comparisons during $[0, t^*]$ and treatment comparisons in $(t^*, \max_j T_j]$. The log hazard ratios are estimated by $\tilde{\boldsymbol{b}} = (\tilde{\boldsymbol{Z}}' \boldsymbol{V}_{\tilde{\boldsymbol{d}}}^{-1} \tilde{\boldsymbol{Z}})^{-1} \tilde{\boldsymbol{Z}}' \boldsymbol{V}_{\tilde{\boldsymbol{d}}}^{-1} \tilde{\boldsymbol{d}}$ and a consistent estimator of the covariance matrix of $\tilde{\boldsymbol{b}}$ is $\boldsymbol{V}_{\tilde{\boldsymbol{b}}} = (\tilde{\boldsymbol{Z}}' \boldsymbol{V}_{\tilde{\boldsymbol{d}}}^{-1} \tilde{\boldsymbol{Z}})^{-1}$. Individual hypothesis tests for each log hazard ratio may be conducted through the statistic $\tilde{b}_{im}^2 / v_{\tilde{b},im}$ where $\tilde{b}_{im}$ is the estimated log hazard ratio for the ith treatment comparison in the mth time interval ($m = 1, 2$), and $v_{\tilde{b},im}$ is its estimated variance. This statistic has an approximate chi-squared distribution with one degree of freedom. A $100(1-\alpha)\%$ confidence interval for each log hazard ratio may be obtained through $\tilde{b}_{im} \pm z_{1-\alpha/2} \sqrt{v_{\tilde{b},im}}$. As in Section 4.2.1, random imbalances in the means of the covariables may be assessed with the statistic $\tilde{Q}_0 = (\tilde{\boldsymbol{d}} - \tilde{\boldsymbol{Z}}\tilde{\boldsymbol{b}})' \boldsymbol{V}_{\tilde{\boldsymbol{d}}}^{-1} (\tilde{\boldsymbol{d}} - \tilde{\boldsymbol{Z}}\tilde{\boldsymbol{b}})$ which has an approximate chi-squared distribution with $P(s-1)$ degrees of freedom.

4.3 Example

The clinical trial data for management of an incurable neurologic disorder from Examples 3.3.1 and 3.3.2 were re-examined for this example. For the subsequent analyses, all 959 patients randomized to one of four treatment groups (placebo, 50 mg, 100 mg, and 200 mg) were included. The primary treatment comparison for the study was 100 mg vs. placebo. The same six baseline covariates of age, disease duration (years), weight, and three continuous measures of neurologic and musculoskeletal function as the analysis in Examples 3.3.1 and 3.3.2 were used. However, the following examples do not include stratification by geographic region/site of disease onset. The primary event of interest was time until progression of disease.

4.3.1 Comparison of Four Treatment Groups: Neurologic Disorder Data

Time to progression of disease is described via Kaplan-Meier curves in Figure 4.1. The curves appear to overlap for about the first 100 days after randomization before they begin to separate. The treatment groups then all appear to have better progression-free survival than the placebo group for the remainder of the observation time. There are times where the 100 mg group appears to perform better than the 200 mg group, although the two curves do cross over each other a few times past the 100-day mark. The 50 mg and 100 mg curves appear to have clear separation, but the 50 mg and 200 mg curves do eventually cross at about 435 days.

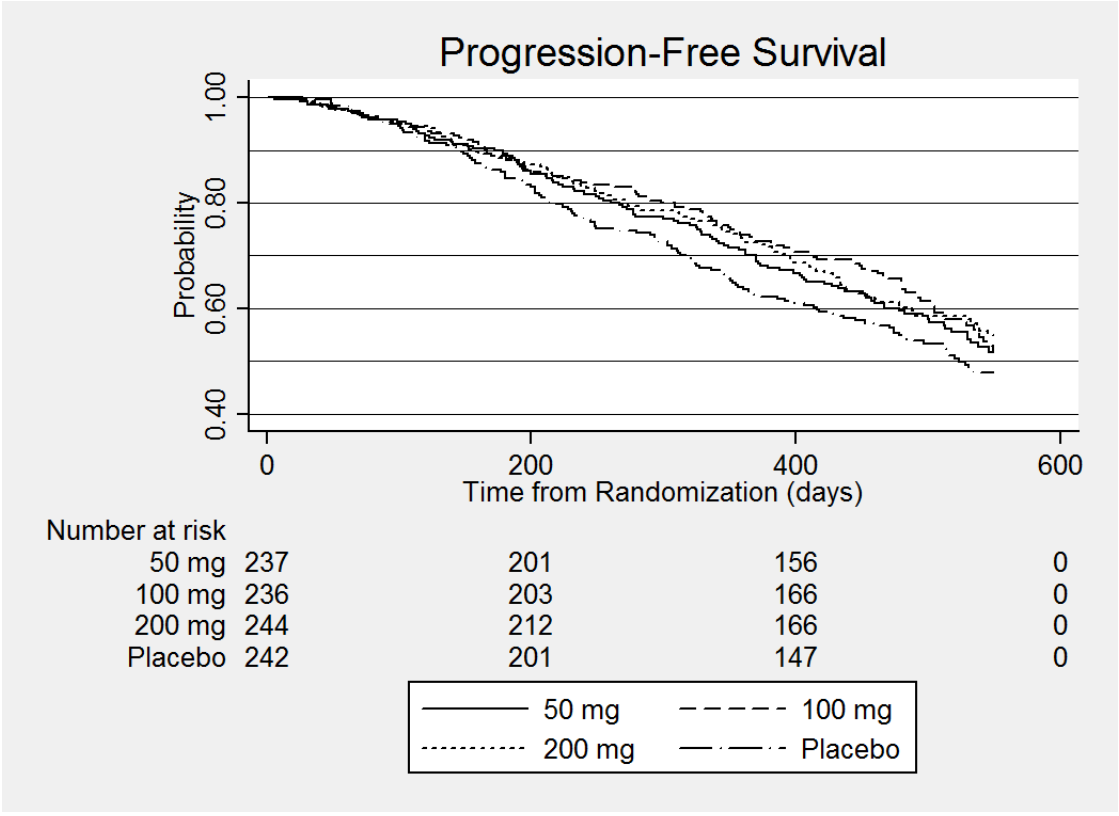

Figure 4.1: Progression-Free Survival for Neurologic Data from Example 4.3.1 with Four Treatment Groups

The results for the treatment comparisons with placebo are presented in Table 4.1. The unadjusted analyses did not show any statistically significant treatment effects (at the 0.05 level) for any of the treatment groups as compared to placebo. The p-values for these comparisons were p=0.2545, p=0.0864, and p=0.0950 for the 50 mg, 100 mg, and 200 mg treatment groups, respectively. After adjustment for baseline covariates via NPANCOVA, the 50 mg vs. placebo comparison showed a 16.5% reduction in the standard error as compared to the unadjusted analysis (NPANCOVA S.E.=0.1121 vs. unadjusted S.E.=0.1342). A 16.7% reduction in standard error due to adjustment via NPANCOVA was observed for both the 100 mg vs. placebo and the 200 mg vs. placebo comparisons. Statistical significance at the 0.05 level was attained for these latter two adjusted comparisons (p=0.0065) and (p=0.0084).

For all three treatment comparisons, the Cox proportional hazards model produced more significant p-values than NPANCOVA. The p-values for the Cox model were due to the estimates being further away from the null value of zero and having larger standard errors than both the NPANCOVA estimates and the unadjusted estimates. The NPANCOVA hazard ratios were closer to 1.0 than the hazard ratios from the Cox model since the NPANCOVA estimates are like population-averaged hazard ratios and the Cox hazard ratios have a conditional interpretation through individuals from different treatment groups who share the same values of the covariates. The assessment of random imbalances in covariate means between the treatment groups was supportive ($Q_0 = 9.2352, df = 18, p = 0.8649$) and indicated a lack of evidence for random imbalances.

Following the guidance of Tangen and Koch (2001) and Kong *et. al.* (2005), a test of $H_0 : \boldsymbol{c}'\boldsymbol{b} = 0$ for the average of 200 mg and 100 mg doses versus placebo was conducted using the contrast $\boldsymbol{c}' = (0\ 1\ 1)$ since the ordering of the estimates in $\boldsymbol{b}$ are

50 mg vs. placebo, 100 mg vs. placebo, and 200 mg vs. placebo. This global assessment had p=0.0970 unadjusted and p=0.0117 adjusted, indicating a treatment effect over placebo for the covariate-adjusted estimates. Subsequent tests of 200 mg versus placebo and 100 mg versus placebo (for covariate-adjusted estimates) were statistically significant with p=0.0065 and 0.0084 as described above. Given that these were both significant, one could test 50 mg versus placebo (p=0.1229). It is at this point that testing would stop and one would conclude efficacy of both the 100 mg and 200 mg doses without formal comparison of these two doses. Alternatively, one could examine the test for homogeneity which had p=0.2735 unadjusted and p=0.0180 adjusted, indicating covariate-adjusted treatment effects were different for at least one pair of dose groups.

4.3.2 Comparison of Four Treatment Groups with Time-Varying Treatment Effects

The data from the previous analysis was considered with the inclusion of time-varying treatment effects. The treatment effect might possibly delay progression of the neurologic disorder for a period of time but not postpone it indefinitely. Thus, separate hazard ratios were of interest for an earlier period (0-12 months) and a later period (12-18 months). The implementation of the methodology in Section 4.2.2 involved the creation of an indicator variable $w(t)$ where $w(t) = 1$ if $t > 365.25$ (days) and $w(t) = 0$ if $t \leq 365.25$ (days).

Estimates of the parameters $(\gamma_1, \gamma_2, \gamma_3)$ for the three interaction terms are presented in Table 4.2. The Cox model analysis with adjustment for the six baseline covariates did not show evidence of interaction for any of the treatment comparisons (p>0.20 for all comparisons). For the unadjusted analysis, there was weak evidence of differences in the log hazard ratios for 0-12 months and 12-18 months for the 50 mg vs. placebo

(p=0.1781) and the 200 mg vs. placebo (p=0.1285) comparisons. However, the 100 mg vs. placebo comparison (p=0.0485) suggested the presence of interaction. After adjustment for covariates, the NPANCOVA analysis corroborated the results of the unadjusted analysis with p=0.1533, p=0.0531, and p=0.1406 for the 50 mg vs. placebo, 100 mg vs. placebo, and 200 mg vs. placebo comparisons, respectively.

The estimated log hazard ratios for the three treatment comparisons for 0-12 months and 12-18 months are presented in Table 4.3. For the 0-12 month 50 mg vs. placebo, 12-18 month 100 mg vs. placebo, and 12-18 month 200 mg vs. placebo comparisons, the NPANCOVA analysis produced covariate-adjusted estimates that were between the unadjusted estimates and the Cox adjusted estimates. For the other comparisons, the NPANCOVA analysis produced covariate-adjusted estimates that were slightly larger than the unadjusted estimates. However, the hazard ratio estimates for the NPANCOVA analysis were similar to those from the unadjusted estimates, reflecting the unconditional nature of the NPANCOVA estimates. This is in contrast to the estimated hazard ratios from the Cox model which were further away from the null value of 1.0 and which have a conditional interpretation on the basis of individuals with the same values for the baseline covariates.

For the 0-12 month time period, there were small reductions in the standard error after covariate adjustment via NPANCOVA (vs. unadjusted) for all three treatment comparisons. Both the unadjusted and NPANCOVA analyses suggest a significant effect (at the 0.05 level) of the 100 mg dose over placebo (p=0.0128 and p=0.0157) as well as the 200 mg dose over placebo (p=0.0257 and p=0.0251), but there was not enough evidence to conclude the 50 mg dose was efficacious (p=0.0918 and p=0.0804). For this time period, the effect of the 100 mg dose was observed to be slightly stronger than the effect of the 200 mg dose, yet the two groups had similar effects. Although the standard errors for the estimates from the Cox model were larger than the NPANCOVA

analyses (with the exception of the 50 mg vs. placebo comparison, the estimates themselves were also of greater magnitude and produced significant p-values for all three treatment comparisons.

For the 12-18 month time period, small reductions in the standard errors were again observed for NPANCOVA as compared to the unadjusted analysis, but none of the estimated treatment effects were statistically significant. The unadjusted and NPANCOVA analyses produced estimated hazard ratios which favored placebo over the treatment, but the Cox model estimates continued to produce estimated hazard ratios less than 1.0, indicating the possible influence of Simpson's paradox (Simpson, 1951).

4.4 Discussion

This chapter presents nonparametric methodology for obtaining covariate-adjusted estimates of log hazard ratios when individuals are randomized to one of several treatment groups. These methods expand upon work by Saville and Koch for two randomized groups as an alternative to the Cox proportional hazards model (Saville and Koch, 2013). The NPANCOVA methods force the differences in means of covariables between treatment groups to zero as is expected under a valid randomization, thereby obtaining the covariate-adjusted treatment effects. The main assumptions for the NPANCOVA method are a valid randomization, adequate sample size to invoke approximate normality of the estimated log hazard ratios, and the proportional hazards assumption on the treatment variables. However, this latter assumption may be avoided by estimating separate covariate-adjusted log hazard ratios in time intervals where the hazard ratios are assumed constant.

The NPANCOVA methods do not require the assumption of proportional hazards for the covariates, and thus they are an attractive alternative to Cox proportional

hazards models in a regulatory setting where the proportional hazards assumption on the covariates can not be verified during the planning stages of a clinical trial. For the example, one of the six baseline covariates showed violation of the proportional hazards assumption, causing possible concern about the Cox model estimates. Additionally, the methods presented here generally produce smaller standard errors of the log hazard ratios as compared to the unadjusted analyses. This is manifested in smaller p-values and thus more powerful tests.

Since the methods in this chapter require the specification of a referent treatment group, the pairwise comparisons between each treatment group and the referent group are correlated. Instead of considering separate NPANCOVA analyses for each treatment comparison, the methods presented here account for this correlation. The methodology can also accommodate contrasts of the adjusted log hazard ratio estimates. For the example, such contrasts could produce the log hazard ratios for the pairwise comparisons between the 50 mg, 100 mg, and 200 mg treatment groups after a global assessment of efficacy through a test of linear trend.

There are limitations associated with the NPANCOVA methods presented here. The covariate adjustment via weighted least squares requires complete covariate data. In the event of missing covariate data, an appropriate multiple imputation strategy would be warranted so as not to potentially bias treatment effects via removal of patients with missing data from the analysis. Removing patients may affect the distribution of covariables and thus the expected differences in means of the covariables between treatment groups may no longer be zero. Additionally, the NPANCOVA methods produce population-averaged estimates of treatment effects and hence can not produce estimated log hazard ratios in subgroups defined by the covariates. Additionally, interactions between treatment variables and covariates can not be accommodated. The interaction of the treatment variable with continuous time in a Cox model is a convenient way to

accommodate a time-varying treatment effect. However, time-varying treatment effects in NPANCOVA may only be handled through the estimation of separate hazard ratios for pre-determined time intervals.

The NPANCOVA methods here are appropriate in a regulatory environment to provide covariate-adjusted estimates of treatment effects for the primary analysis in a clinical trial. The methods avoid the proportional hazards assumption on the covariates, and thus the authors recommend the use of the NPANCOVA methods for the primary analysis followed by supportive secondary analyses using the Cox model. These secondary analyses could explore subgroup differences in treatment effects, and the effects of model assumptions on the results could be assessed through comparison with the NPANCOVA results.

Table 4.1: Results from Analysis on Four Treatment Groups with No Time-Varying Treatment

Comparison	Method	Log HR	S.E.	HR	95% CI		p-value
					Lower	Upper	
50 mg vs. placebo	Unadjusted	-0.1529	0.1342	0.8582	0.6598	1.1164	0.2545
	NPANCOVA	-0.1730	0.1121	0.8411	0.6752	1.0479	0.1229
	Cox Adjusted	-0.2850	0.1339	0.7520	0.5784	0.9777	0.0333
100 mg vs. placebo	Unadjusted	-0.2299	0.1341	0.7946	0.6110	1.0335	0.0864
	NPANCOVA	-0.3037	0.1117	0.7381	0.5930	0.9186	0.0065
	Cox Adjusted	-0.4327	0.1362	0.6488	0.4968	0.8473	0.0015
200 mg vs. placebo	Unadjusted	-0.2261	0.1354	0.7976	0.6117	1.0401	0.0950
	NPANCOVA	-0.2975	0.1129	0.7427	0.5952	0.9267	0.0084
	Cox Adjusted	-0.5148	0.1371	0.5976	0.4568	0.7819	0.0002

Table 4.2: Parameter Estimates for Interactions from Analysis with Time-Varying Treatment

Comparison	Method	Estimate	S.E.	95% CI		p-value
				Lower	Upper	
50 mg vs. Placebo	Unadjusted	0.3948	0.2932	-0.1799	0.9695	0.1781
	NPANCOVA	0.4149	0.2905	-0.1545	0.9843	0.1533
	Cox Adjusted	0.2691	0.2932	-0.3056	0.8438	0.3588
100 mg vs. Placebo	Unadjusted	0.5705	0.2892	0.0037	1.1373	0.0485
	NPANCOVA	0.5551	0.2870	-0.0074	1.1176	0.0531
	Cox Adjusted	0.3677	0.2920	-0.2046	0.9400	0.2079
200 mg vs. Placebo	Unadjusted	0.4490	0.2954	-0.1299	1.0280	0.1285
	NPANCOVA	0.4320	0.2931	-0.1425	1.0065	0.1406
	Cox Adjusted	0.2182	0.2972	-0.3643	0.8007	0.4628

Table 4.3: Results from Analysis with Time-Varying Treatment

Time Period	Comparison	Method	Log HR	S.E.	HR	95% CI		p-value
						Lower	Upper	
0-12 months	50 mg vs. Placebo	Unadjusted	-0.2676	0.1587	0.7652	0.5606	1.0445	0.0918
		NPANCOVA	-0.2753	0.1575	0.7594	0.5577	1.0339	0.0804
		Cox Adjusted	-0.3604	0.1556	0.6974	0.5141	0.9460	0.0205
	100 mg vs. Placebo	Unadjusted	-0.4104	0.1648	0.6634	0.4802	0.9163	0.0128
		NPANCOVA	-0.3966	0.1641	0.6726	0.4876	0.9279	0.0157
		Cox Adjusted	-0.5471	0.1677	0.5786	0.4165	0.8038	0.0011
	200 mg vs. Placebo	Unadjusted	-0.3593	0.1611	0.6982	0.5092	0.9574	0.0257
		NPANCOVA	-0.3581	0.1599	0.6990	0.5109	0.9563	0.0251
		Cox Adjusted	-0.5727	0.1638	0.5640	0.4091	0.7775	0.0005
12-18 months	50 mg vs. Placebo	Unadjusted	0.1272	0.2466	1.1357	0.7004	1.8414	0.6058
		NPANCOVA	0.1396	0.2447	1.1498	0.7117	1.8576	0.5684
		Cox Adjusted	-0.0913	0.2615	0.9127	0.5467	1.5239	0.7270
	100 mg vs. Placebo	Unadjusted	0.1600	0.2377	1.1735	0.7365	1.8699	0.5008
		NPANCOVA	0.1568	0.2356	1.1718	0.7385	1.8594	0.5009
		Cox Adjusted	-0.1794	0.2469	0.8357	0.5151	1.3559	0.4673
	200 mg vs. Placebo	Unadjusted	0.0898	0.2478	1.0939	0.6731	1.7779	0.7171
		NPANCOVA	0.0739	0.2465	1.0767	0.6641	1.7455	0.7645
		Cox Adjusted	-0.3545	0.2687	0.7015	0.4143	1.1879	0.1871

Chapter 5

Summary and Future Research

5.1 Summary

The methods in the previous chapters describe several extensions of nonparametric randomization-based analysis of covariance. Chapter 2 provides methodology for covariance adjustment in the analysis of dichotomous outcomes for 1:1 matched sets, $M:1$ matched sets, and the 1:1 non-randomized matched case control study. These techniques are an alternative to either the unadjusted McNemar's test or conditional logistic regression. The NPANCOVA methods avoid the assumptions of functional form of the covariates, and they only require a valid randomization (for exact inference). Confidence intervals for the covariate-adjusted difference in proportions (or odds ratio) may be obtained if the study participants can be assumed to be like a simple random sample from a corresponding population. For the examples presented, a decrease in standard errors for the treatment effect estimates was observed, resulting in more precise estimates.

Chapter 3 extends the methodology of Saville and Koch (2012) to the analysis of dichotomous and time-to-event outcomes in the presence of stratification. DFBETA residuals from a treatment-only conditional logistic regression (or stratified Cox proportional hazards model, for the time-to-event outcome) are used to create a robust

estimate of the covariance matrix for the model-based log odds ratio (or log hazard ratio). This information is combined with covariate information to produce a covariance matrix which can be used for the nonparametric covariance adjustment via weighted least squares. The nonparametric covariance adjustment produces standard errors for the estimated treatment effect that are smaller in comparison to the standard errors from a stratified analysis with no other covariates besides treatment.

The methods of Saville and Koch are also extended to the setting of multiple randomized treatment groups for time-to-event outcomes. The techniques presented begin with a treatment-only Cox proportional hazards model and use the DFBETA residuals from this model to obtain a robust covariance matrix for the log hazard ratios. The DFBETA residuals are then combined with the covariate information to obtain a covariance matrix which can be used in the nonparametric covariance adjustment. The difference vector in this setting uses pairwise differences between each treatment group and a referent group, although contrasts for the covariate-adjusted log hazard ratios may be used to assess trends corresponding to increasing dose. Tangen and Koch (2001) provide guidance on how these hypothesis tests can be conducted.

5.2 Considerations and Limitations

5.2.1 Missing Data

For all of the methods above, there is an assumption that none of the outcome data or covariate data is missing. The methods do not have a capability to handle missing data directly. In the presence of missing data in a randomized trial, analyzing only complete cases might result in the two randomized groups not being strictly comparable on the basis of the covariables. Thus, it is recommended that the missing covariate data be imputed by some appropriate multiple imputation method before applying

NPANCOVA. For missing outcome variables, the missing data mechanism (e.g. MCAR, MAR, or NMAR) should be posited, and then appropriate methods considered for handling the missingness. The NPANCOVA methods are appropriate for analysis of the primary endpoint of a study, and NPANCOVA could be applied to imputed data sets in a sensitivity analysis to assess the influence of the missing data.

5.2.2 Limitations

The nonparametric covariance adjustment described for these methods is done via weighted least squares using a model which forces the differences in means of the covariables between the randomized groups to zero. As a result, relationships between the covariables and the outcome can not be estimated. Thus, for the regulatory setting, the NPANCOVA methods are intended for analysis of the primary endpoint in a clincal trial. Once efficacy is established through NPANCOVA, secondary analyses may be conducted using parametric models. These models may include assessment of treatment effects for certain subgroups or interactions between treatment effects and levels of the other covariates.

5.3 Future Research

The methods for matched sets presented in Chapter 2 are appropriate in the case of balanced allocation. Example 2.3.2 does involve the use of weighting estimates across type of matched sets, although the sample sizes within the matched sets are the same in this case. A further direction of research for this methodology may involve the extension to situations of unbalanced allocation when the matched sets have variable size. The extension of the NPANCOVA methodology to randomized studies with M:N matching (where M and N are both ≥ 2 but ≤ 4) is straightforward. The difference in mean responses would be formed for the two treatment groups for each matched

set, and formation of $\bar{\boldsymbol{f}}$ and $\boldsymbol{V}_{\bar{\boldsymbol{f}}}$ would proceed as previously outlined. For the odds ratio, the methodology presented in Section 2.2.3 could be extended via a Mantel-Haenszel estimate of the odds ratio or a conditional maximum likelihood estimate as in (Miettinen, 1970), but this is beyond the scope of this research. In practice, M and N are rarely greater than 3, and so the methodology here is most applicable for matched pairs and matched sets of size ≤ 6. For larger matched sets, the methodology presented in Chapter 3 may be more appropriate.

For the NPANCOVA methodology as applied to matched sets in Chapter 2, several extensions of the NPANCOVA methdology are possible. While only dichotomous outcomes are considered, these methods may be easily extended to continuous outcomes through a covariate-adjusted difference in means. Repeated outcome measures through multiple assessments (e.g. clinic visits) could be managed by including entries in the difference vector for each pair corresponding to the differences that the respective assessments have for the responses for the two treatments. In the case of 1:1:1 randomized allocation to treatments A,B, or C (matched triples), the difference vector for stratum h could contain the A-C difference in response, the B-C difference in response, and such differences for the covariates for adjustment. Formation of the appropriate mean vector and its covariance matrix would follow per Section 2.2.1. Subsequent assessment of ordinality of the treatment effects could be accomplished via hypothesis testing of contrasts for the vector of covariate-adjusted estimates.

The methods in Chapter 3 and 4 could be combined in the case of a randomized study with stratification and multiple treatment groups. Extension to ordinal outcomes and counts is a possibility, although it has not yet been determined whether the cross-products of the DFBETA residuals from a treatment-only proportional odds model or Poisson regression model can be used to obtain a valid robust estimate of the covariance matrix for the log odds ratios or log rate ratios. In the case of repeated measures, there

is the possible extension of the Saville and Koch methodology using the DFBETA residuals from generalized estimating equations (GEE). Preisser and Qaqish (1996) have developed cluster-level DFBETA residuals (i.e. DFBETA residuals based on deleting an entire cluster's observations) and observation-level DFBETA residuals (i.e. based on deleting one observation within a cluster) which are available in standard software, and the cross-products of the cluster-level DFBETA residuals have been shown to exactly equal a bias-corrected covariance estimator in the GEE setting (Mancl and Derouen, 2001). Accommodating baseline covariates in this setting is straightforward, and it would be possible to use two or more of the repeated measures (e.g. a pre-baseline measure and its baseline counterpart) as covariates. However, it is currently unclear how to easily manage adjustment of similar measures on different scales (e.g. a dichotomized baseline value included as an outcome measure and its continuous counterpart included as a baseline covariate).

While the research presented here offers a theoretical road map for the NPANCOVA methods in these settings, a future direction of the research could be to assess the performance of the NPANCOVA methods in this paper as they relate to the parametric modeling methods. Simulations would be a necessary next step to ensure that the NPANCOVA methods do not result in any Type I error increase beyond the nominal level as well as to ensure that the methods have adequate power for testing the hypotheses of interest. From the examples presented here, one can observe that the power to test coefficients relating to the treatment effects is greater when there is covariance adjustment than when there is no covariance adjustment (when the standard errors show a decrease after adjustment). However, regarding power, it is unclear how the NPANCOVA methods perform compared to the parametric ANCOVA methods. Jiang *et. al.* have conducted a simulation study comparing the NPANCOVA methods of Tangen and Koch (1999) for logrank and Wilcoxon scores to the Cox proportional

hazards model. They showed that the NPANCOVA methods did preserve the nominal Type I error and that there was a small loss in efficiency of the NPANCOVA methods when the proportional hazards assumption was correct (Jiang et al., 2008). However, in planning a study, the cost of having a greater sample size for NPANCOVA may be outweighed by the benefit of having minimal statistical assumptions when pre-specifying the primary analysis method. Accordingly, the sample size needed for NPANCOVA may end up being less than what would be necessary in a nonparametric analysis that did not adjust for covariables (e.g. logrank test).

Appendix A: Chapter 3

A.1 Specification of Covariance Matrix $\tilde{V}_f$ for Compound Vector in Stratified Analysis using Multivariate U-statistics

A.1.1 Theoretical Specification

Methods for multivariate U-statistics ((Davis and Quade, 1968),(Davis and Quade, 1978),(Kawaguchi, Koch and Wang, 2011)) are used to construct consistent estimates (different from that in (3.3)) for the covariances of the log odds ratio $\hat{\beta}$ and the stratified difference in means of the covariates $(\bar{\boldsymbol{x}}_{\mathbf{1}} - \bar{\boldsymbol{x}}_{\mathbf{2}})$ as well as for the covariance matrix of $(\bar{\boldsymbol{x}}_{\mathbf{1}} - \bar{\boldsymbol{x}}_{\mathbf{2}})$. Let $\boldsymbol{g}_{jj'} = 0.5 * [I(S_j - S_{j'} = 0)(t_j - t_{j'})(\boldsymbol{x}_j - \boldsymbol{x}_{j'})]n(n-1)/(2W(n_j + n_{j'}))$, where $n_j = n_{hi}$ if patient j is from group i in stratum h (with n_{hi} being the sample size for group i in stratum h), S_j denotes the stratum for patient j, $t_j = 1$ if patient j is from group 1 and $t_j = -1$ if patient j is from group 2, and $\boldsymbol{x}_j$ is the vector of covariates for patient j. Also, $I(S_j - S_{j'} = 0) = 1$ if $S_j = S_{j'}$ and it equals 0 if $S_j \neq S_{j'}$; Additionally, let $W = \sum_{h=1}^{q}[n_{h1}n_{h2}/(n_h)]$ where q is the number of strata with $n_h = n_{h1} + n_{h2}$; also let $n = \sum_{h=1}^{q}\sum_{i=1}^{2} n_{hi}$. Let $\boldsymbol{g}_{j*}$ be the mean of the $\boldsymbol{g}_{jj'}$ for patient j with averaging across all $j' \neq j$, so that the mean $\bar{\boldsymbol{g}}$ of the $\boldsymbol{g}_{j*}$ across all patients equals the weighted difference in baseline covariate means between the two groups for the respective strata (with weights $(n_{h1}n_{h2}/n_h)$) with this being denoted as the stratified difference $\bar{\boldsymbol{g}} = (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)$.

$$\boldsymbol{g}_{j*} = \sum_{j \neq j'}^{n} \boldsymbol{g}_{jj'}/(n-1) = \begin{cases} \frac{nn_{h2}(\boldsymbol{x}_j - \bar{\boldsymbol{x}}_{h2})}{2W(n_{h1}+n_{h2})} & \text{if } j \text{ from } h, \text{ group 1} \\ \frac{-nn_{h1}(\boldsymbol{x}_j - \bar{\boldsymbol{x}}_{h1})}{2W(n_{h1}+n_{h2})} & \text{if } j \text{ from } h, \text{ group 2} \end{cases}$$

$$\bar{\boldsymbol{g}} = \sum_{j=1}^{n} \boldsymbol{g}_{j*}/n = \left[\sum_{h=1}^{q} \frac{n_{h1}n_{h2}(\bar{\boldsymbol{x}}_{h1} - \bar{\boldsymbol{x}}_{h2})}{(n_{h1}+n_{h2})} \right] /W = \bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2 \tag{1}$$

The quantity $(\boldsymbol{g}_{j*} - \bar{\boldsymbol{g}})$ as given is a measure of the deviation of $\boldsymbol{g}_{j*}$ about its mean:

$$(\boldsymbol{g}_{j*} - \bar{\boldsymbol{g}}) = \begin{cases} \frac{nn_{h2}(\boldsymbol{x}_j - \bar{\boldsymbol{x}}_{h2})}{2W(n_{h1}+n_{h2})} - (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2) & \text{if } j \text{ from } h, \text{ group 1} \\ \frac{-nn_{h1}(\boldsymbol{x}_j - \bar{\boldsymbol{x}}_{h1})}{2W(n_{h1}+n_{h2})} - (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2) & \text{if } j \text{ from } h, \text{ group 2} \end{cases}$$

$\tilde{\boldsymbol{V}}_{\bar{\boldsymbol{g}}}$ in (2) is a consistent estimate for the covariance matrix of $\bar{\boldsymbol{g}}$ ((Davis and Quade, 1968),(Davis and Quade, 1978),(Kawaguchi, Koch and Wang, 2011)).

$$\tilde{\boldsymbol{V}}_{\bar{\boldsymbol{g}}} = \frac{4}{n(n-1)} \sum_{j=1}^{n} (\boldsymbol{g}_{j*} - \bar{\boldsymbol{g}})(\boldsymbol{g}_{j*} - \bar{\boldsymbol{g}})' = \sum_{j=1}^{n} \tilde{\boldsymbol{g}}_j \tilde{\boldsymbol{g}}_j' \tag{2}$$

The quantity $\tilde{\boldsymbol{g}}_j$ in (2) is defined as

$$\tilde{\boldsymbol{g}}_j = \frac{2}{\sqrt{n(n-1)}} (\boldsymbol{g}_{j*} - \bar{\boldsymbol{g}})$$

$$= \begin{cases} \sqrt{\frac{n}{n-1}} \left[\frac{n_{h2}(\boldsymbol{x}_j - \bar{\boldsymbol{x}}_{h2})}{W(n_{h1}+n_{h2})} \right] - \frac{2(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)}{\sqrt{n(n-1)}} & \text{if } j \text{ from } h, \text{ group 1} \\ -\sqrt{\frac{n}{n-1}} \left[\frac{n_{h1}(\boldsymbol{x}_j - \bar{\boldsymbol{x}}_{h1})}{W(n_{h1}+n_{h2})} \right] - \frac{2(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)}{\sqrt{n(n-1)}} & \text{if } j \text{ from } h, \text{ group 2} \end{cases}$$

$$= \begin{cases} \sqrt{\frac{n}{n-1}} \left[\frac{n_{h2}(\boldsymbol{x}_j - \bar{\boldsymbol{x}}_{h1})}{W(n_{h1}+n_{h2})} \right] + \sqrt{\frac{n}{n-1}} \left[\frac{n_{h2}(\bar{\boldsymbol{x}}_{h1} - \bar{\boldsymbol{x}}_{h2})}{W(n_{h1}+n_{h2})} \right] - \frac{2(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)}{\sqrt{n(n-1)}} & \text{if } j \text{ from } h, \text{ group 1} \\ \sqrt{\frac{n}{n-1}} \left[\frac{-n_{h1}(\boldsymbol{x}_j - \bar{\boldsymbol{x}}_{h2})}{W(n_{h1}+n_{h2})} \right] + \sqrt{\frac{n}{n-1}} \left[\frac{n_{h1}(\bar{\boldsymbol{x}}_{h1} - \bar{\boldsymbol{x}}_{h2})}{W(n_{h1}+n_{h2})} \right] - \frac{2(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)}{\sqrt{n(n-1)}} & \text{if } j \text{ from } h, \text{ group 2} \end{cases}$$

Let $\tilde{\boldsymbol{f}}_j = (r_j, \tilde{\boldsymbol{g}}_j')'$ and $\boldsymbol{f} = (\hat{\beta}, (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)')'$, where r_j is the DFBETA residual for the jth patient. Then the covariance matrix of $\boldsymbol{f}$ is estimated by (3), for which the lower right quantity $\sum_{j=1}^{n} \tilde{\boldsymbol{g}}_j \tilde{\boldsymbol{g}}_j'$ estimates the covariance matrix of the weighted differences

$\bar{\boldsymbol{g}} = (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)$ in means of covariates, and the upper right quantity $\sum_{j=1}^{n} r_j \tilde{\boldsymbol{g}}'_j$ estimates the covariances of the estimated treatment effect $\hat{\beta}$ and the weighted differences in means of covariates.

$$\begin{aligned}
\tilde{\boldsymbol{V}}_{\boldsymbol{f}} &= \sum_j \tilde{\boldsymbol{f}}_j \tilde{\boldsymbol{f}}'_j \\
&= \begin{bmatrix} \sum_{j=1}^{n} r_j^2 & \sum_{j=1}^{n} r_j \tilde{\boldsymbol{g}}'_j \\ \sum_{j=1}^{n} \tilde{\boldsymbol{g}}_j r_j & \sum_{j=1}^{n} \tilde{\boldsymbol{g}}_j \tilde{\boldsymbol{g}}'_j \end{bmatrix} \\
&= \begin{bmatrix} v_{\hat{\beta}} & \tilde{\boldsymbol{V}}'_{\hat{\beta},\bar{\boldsymbol{g}}} \\ \tilde{\boldsymbol{V}}_{\hat{\beta},\bar{\boldsymbol{g}}} & \tilde{\boldsymbol{V}}_{\bar{\boldsymbol{g}}} \end{bmatrix}
\end{aligned} \tag{3}$$

The quantities in $\tilde{\boldsymbol{V}}_{\boldsymbol{f}}$ are estimated by

$$\begin{aligned}
\sum_{j=1}^{n} \tilde{\boldsymbol{g}}_j \tilde{\boldsymbol{g}}'_j = \sum_{h=1}^{q} \sum_{i=1}^{2} \sum_{k=1}^{n_{hi}} \frac{n(n_{hi}-1)}{(n-1)n_{hi}} \left[\frac{n_{h1}n_{h2}}{n_h W}\right]^2 \left[\frac{(\boldsymbol{x}_{hik} - \bar{\boldsymbol{x}}_{hi})(\boldsymbol{x}_{hik} - \bar{\boldsymbol{x}}_{hi})'}{n_{hi}(n_{hi}-1)}\right] \\
+ \sum_{h=1}^{q} \frac{n}{n-1} \frac{n_{h1}n_{h2}}{n_h W^2} (\bar{\boldsymbol{x}}_{h1} - \bar{\boldsymbol{x}}_{h2})(\bar{\boldsymbol{x}}_{h1} - \bar{\boldsymbol{x}}_{h2})' \\
- \frac{4}{n-1} (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)'
\end{aligned}$$

$$\begin{aligned}
\sum_{j=1}^{n} r_j \tilde{\boldsymbol{g}}'_j = \sqrt{\frac{n}{n-1}} \sum_{h=1}^{q} \left(\frac{1}{W n_h}\right) \left[n_{h2} \sum_{k=1}^{n_{h1}} r_{h1k} (\boldsymbol{x}_{h1k} - \bar{\boldsymbol{x}}_{h1}) - n_{h1} \sum_{k=1}^{n_{h2}} r_{h2k} (\boldsymbol{x}_{h2k} - \bar{\boldsymbol{x}}_{h2})\right] \\
+ \sqrt{\frac{n}{n-1}} \sum_{h=1}^{q} \left(\frac{n_{h1}n_{h2}}{W n_h}\right) (\bar{\boldsymbol{x}}_{h1} - \bar{\boldsymbol{x}}_{h2})(\bar{r}_{h1} + \bar{r}_{h2}) \\
- \frac{2}{\sqrt{n(n-1)}} (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2) \sum_{h=1}^{q} (n_{h1}\bar{r}_{h1} + n_{h2}\bar{r}_{h2})
\end{aligned}$$

Suppose the case of proportional allocation within strata so that $n_{h1} = pn_h$ and $n_{h2} = (1-p)n_h$ for $0 < p < 1$. Then $(n_{h1}n_{h2}/n_h) = p(1-p)n_h$ and $W = p(1-p)n$ so that

$(n_{h1}n_{h2})/(n_hW) = (n_h)/n$. Then,

$$\sum_{j=1}^{n} \tilde{\boldsymbol{g}}_j \tilde{\boldsymbol{g}}_j' = \frac{n}{n-1} \sum_{h=1}^{q} \left[\frac{\sum_{k=1}^{n_{h1}} (\boldsymbol{x}_{h1k} - \bar{\boldsymbol{x}}_{h1})(\boldsymbol{x}_{h1k} - \bar{\boldsymbol{x}}_{h1})'}{n^2 p^2} + \frac{\sum_{k=1}^{n_{h2}} (\boldsymbol{x}_{h2k} - \bar{\boldsymbol{x}}_{h2})(\boldsymbol{x}_{h2k} - \bar{\boldsymbol{x}}_{h2})'}{n^2 (1-p)^2} \right]$$
$$+ \frac{1}{(n-1)p(1-p)} \sum_{h=1}^{q} \left(\frac{n_h}{n} \right) (\bar{\boldsymbol{x}}_{h1} - \bar{\boldsymbol{x}}_{h2})(\bar{\boldsymbol{x}}_{h1} - \bar{\boldsymbol{x}}_{h2})'$$
$$- \frac{4}{n-1} (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)'$$

and

$$\sum_{j=1}^{n} r_j \tilde{\boldsymbol{g}}_j' = \sqrt{\frac{n}{n-1}} \sum_{h=1}^{q} \left[\sum_{k=1}^{n_{h1}} \frac{1}{np} r_{h1k} (\boldsymbol{x}_{h1k} - \bar{\boldsymbol{x}}_{h1}) - \sum_{k=1}^{n_{h2}} \frac{1}{n(1-p)} r_{h2k} (\boldsymbol{x}_{h2k} - \bar{\boldsymbol{x}}_{h2}) \right]$$
$$+ \sqrt{\frac{n}{n-1}} \sum_{h=1}^{q} \left(\frac{n_h}{n} \right) (\bar{\boldsymbol{x}}_{h1} - \bar{\boldsymbol{x}}_{h2})(\bar{r}_{h1} + \bar{r}_{h2})$$
$$- 2\sqrt{\frac{n}{n-1}} (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2) \sum_{h=1}^{q} (p\bar{r}_{h1} + (1-p)\bar{r}_{h2}) \left(\frac{n_h}{n} \right)$$

Furthermore, when stratum sizes are equal so that $n_h = N$ for all h, $n = Nq$, and $(n_h/n) = 1/q$,

$$\sum_{j=1}^{n} \tilde{\boldsymbol{g}}_j \tilde{\boldsymbol{g}}_j' = \frac{Nq}{Nq-1} \sum_{h=1}^{q} \left[\frac{\sum_{k=1}^{Np} (\boldsymbol{x}_{h1k} - \bar{\boldsymbol{x}}_{h1})(\boldsymbol{x}_{h1k} - \bar{\boldsymbol{x}}_{h1})'}{N^2 q^2 p^2} + \frac{\sum_{k=1}^{N(1-p)} (\boldsymbol{x}_{h2k} - \bar{\boldsymbol{x}}_{h2})(\boldsymbol{x}_{h2k} - \bar{\boldsymbol{x}}_{h2})'}{N^2 q^2 (1-p)^2} \right]$$
$$+ \frac{1}{(Nq-1)qp(1-p)} \sum_{h=1}^{q} (\bar{\boldsymbol{x}}_{h1} - \bar{\boldsymbol{x}}_{h2})(\bar{\boldsymbol{x}}_{h1} - \bar{\boldsymbol{x}}_{h2})'$$
$$- \frac{4}{Nq-1} (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)'$$

$$\sum_{j=1}^{n} r_j \tilde{\boldsymbol{g}}_j' = \sqrt{\frac{Nq}{Nq-1}} \sum_{h=1}^{q} \left[\sum_{k=1}^{Np} \frac{1}{Npq} r_{h1k}(\boldsymbol{x}_{h1k} - \bar{\boldsymbol{x}}_{h1}) - \sum_{k=1}^{N(1-p)} \frac{1}{N(1-p)q} r_{h2k}(\boldsymbol{x}_{h2k} - \bar{\boldsymbol{x}}_{h2}) \right]$$
$$+ \sqrt{\frac{Nq}{Nq-1}} \sum_{h=1}^{q} \left(\frac{1}{q}\right) (\bar{\boldsymbol{x}}_{h1} - \bar{\boldsymbol{x}}_{h2})(\bar{r}_{h1} + \bar{r}_{h2})$$
$$- 2\sqrt{\frac{Nq}{Nq-1}} \left(\frac{1}{q}\right) (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2) \sum_{h=1}^{q} (p\bar{r}_{h1} + (1-p)\bar{r}_{h2})$$

In the special case of equal stratum sizes and equal allocation within strata so that $p = 0.5$,

$$\sum_{j=1}^{n} \tilde{\boldsymbol{g}}_j \tilde{\boldsymbol{g}}_j' = \sum_{h=1}^{q} \frac{4Nq}{N^2q^2(Nq-1)} \left[\sum_{i=1}^{2} \sum_{k=1}^{N/2} (\boldsymbol{x}_{hik} - \bar{\boldsymbol{x}}_{hi})(\boldsymbol{x}_{hik} - \bar{\boldsymbol{x}}_{hi})' \right]$$
$$+ \frac{4}{q(Nq-1)} \sum_{h=1}^{q} (\bar{\boldsymbol{x}}_{h1} - \bar{\boldsymbol{x}}_{h2})(\bar{\boldsymbol{x}}_{h1} - \bar{\boldsymbol{x}}_{h2})'$$
$$- \frac{4}{Nq-1} (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)'$$

$$\sum_{j=1}^{n} r_j \tilde{\boldsymbol{g}}_j' = \sqrt{\frac{Nq}{Nq-1}} \left(\frac{2}{Nq}\right) \sum_{h=1}^{q} \left[\sum_{k=1}^{N/2} r_{h1k}(\boldsymbol{x}_{h1k} - \bar{\boldsymbol{x}}_{h1}) - \sum_{k=1}^{N/2} r_{h2k}(\boldsymbol{x}_{h2k} - \bar{\boldsymbol{x}}_{h2}) \right]$$
$$+ \sqrt{\frac{Nq}{Nq-1}} \left(\frac{1}{q}\right) \sum_{h=1}^{q} (\bar{\boldsymbol{x}}_{h1} - \bar{\boldsymbol{x}}_{h2})(\bar{r}_{h1} + \bar{r}_{h2})$$
$$- \sqrt{\frac{Nq}{Nq-1}} (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2) \sum_{h=1}^{q} \frac{\bar{r}_{h1} + \bar{r}_{h2}}{q}$$

A.1.2 Numerical Assessment

In this section, numerical comparisons are made between the covariance matrix presented in Chapter 3, $\boldsymbol{V_f}$, and the covariance matrix presented in Appendix A.1.1, $\tilde{\boldsymbol{V}}_{\boldsymbol{f}}$. Specifically, the components of $\tilde{\boldsymbol{V}}_{\boldsymbol{f}}$ are presented for comparison. The time-to-event outcome for the neurologic disorder data (Example 3.3.2) and the dichotomous outcome for the osteoarthritis data (Example 3.3.3) are used as examples.

Time-to-Event Outcome: Neurologic Disorder Data

Per the methodology presented in Appendix A.1.1, the submatrix $\tilde{\boldsymbol{V}}_{\bar{\boldsymbol{g}}}$ of the covariance matrix $\tilde{\boldsymbol{V}}_{\boldsymbol{f}}$ can be written as the sum of three terms $\tilde{\boldsymbol{V}}_{\bar{\boldsymbol{g}}}^{(1)}$, $\tilde{\boldsymbol{V}}_{\bar{\boldsymbol{g}}}^{(2)}$, and $\tilde{\boldsymbol{V}}_{\bar{\boldsymbol{g}}}^{(3)}$ such that $\tilde{\boldsymbol{V}}_{\bar{\boldsymbol{g}}} = \sum_{j=1}^{n} \tilde{\boldsymbol{g}}_j \tilde{\boldsymbol{g}}_j' = \tilde{\boldsymbol{V}}_{\bar{\boldsymbol{g}}}^{(1)} + \tilde{\boldsymbol{V}}_{\bar{\boldsymbol{g}}}^{(2)} + \tilde{\boldsymbol{V}}_{\bar{\boldsymbol{g}}}^{(3)}$. Here,

$$\tilde{\boldsymbol{V}}_{\bar{\boldsymbol{g}}}^{(1)} = \sum_{h=1}^{q} \sum_{i=1}^{2} \sum_{k=1}^{n_{hi}} \frac{n(n_{hi}-1)}{(n-1)n_{hi}} \left[\frac{n_{h1}n_{h2}}{n_h W}\right]^2 \left[\frac{(\boldsymbol{x}_{hik} - \bar{\boldsymbol{x}}_{hi})(\boldsymbol{x}_{hik} - \bar{\boldsymbol{x}}_{hi})'}{n_{hi}(n_{hi}-1)}\right]$$

$$\tilde{\boldsymbol{V}}_{\bar{\boldsymbol{g}}}^{(2)} = \sum_{h=1}^{q} \frac{n}{n-1} \frac{n_{h1}n_{h2}}{n_h W^2} (\bar{\boldsymbol{x}}_{h1} - \bar{\boldsymbol{x}}_{h2})(\bar{\boldsymbol{x}}_{h1} - \bar{\boldsymbol{x}}_{h2})'$$

$$\tilde{\boldsymbol{V}}_{\bar{\boldsymbol{g}}}^{(3)} = -\frac{4}{n-1} (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)'$$

Similarly, the submatrix $\tilde{\boldsymbol{V}'}_{\hat{\beta},\bar{\boldsymbol{g}}}$ of the covariance matrix $\tilde{\boldsymbol{V}}_{\boldsymbol{f}}$ can be written as the sum of the three terms $\tilde{\boldsymbol{V}}_{\hat{\beta},\bar{\boldsymbol{g}}}^{(1)}$, $\tilde{\boldsymbol{V}}_{\hat{\beta},\bar{\boldsymbol{g}}}^{(2)}$, and $\tilde{\boldsymbol{V}}_{\hat{\beta},\bar{\boldsymbol{g}}}^{(3)}$ such that $\tilde{\boldsymbol{V}'}_{\hat{\beta},\bar{\boldsymbol{g}}} = \sum_{j=1}^{n} r_j \tilde{\boldsymbol{g}}_j' = \tilde{\boldsymbol{V}}_{\hat{\beta},\bar{\boldsymbol{g}}}^{(1)} + \tilde{\boldsymbol{V}}_{\hat{\beta},\bar{\boldsymbol{g}}}^{(2)} + \tilde{\boldsymbol{V}}_{\hat{\beta},\bar{\boldsymbol{g}}}^{(3)}$. Here,

$$\tilde{\boldsymbol{V}}_{\hat{\beta},\bar{\boldsymbol{g}}}^{(1)} = \sqrt{\frac{n}{n-1}} \sum_{h=1}^{q} \left(\frac{1}{W n_h}\right) \left[n_{h2} \sum_{k=1}^{n_{h1}} r_{h1k}(\boldsymbol{x}_{h1k} - \bar{\boldsymbol{x}}_{h1}) - n_{h1} \sum_{k=1}^{n_{h2}} r_{h2k}(\boldsymbol{x}_{h2k} - \bar{\boldsymbol{x}}_{h2})\right]$$

$$\tilde{\boldsymbol{V}}_{\hat{\beta},\bar{\boldsymbol{g}}}^{(2)} = \sqrt{\frac{n}{n-1}} \sum_{h=1}^{q} \left(\frac{n_{h1}n_{h2}}{W n_h}\right) (\bar{\boldsymbol{x}}_{h1} - \bar{\boldsymbol{x}}_{h2})(\bar{r}_{h1} + \bar{r}_{h2})$$

$$\tilde{\boldsymbol{V}}_{\hat{\beta},\bar{\boldsymbol{g}}}^{(3)} = -\frac{2}{\sqrt{n(n-1)}} (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2) \sum_{h=1}^{q} (n_{h1}\bar{r}_{h1} + n_{h2}\bar{r}_{h2})$$

The numerical values of the above components for the neurologic disorder data are

as follows:

$$\tilde{\boldsymbol{V}}_{\bar{\boldsymbol{g}}}^{(1)} = \begin{bmatrix} 0.735674 & 0.007816 & 0.007736 & -0.078093 & -0.128323 & 0.201556 \\ & 1.898633 & -0.018097 & 0.503850 & -0.028688 & -0.359576 \\ & & 0.010596 & -0.023594 & -0.002767 & 0.027834 \\ & & & 1.296853 & 0.136271 & -0.503879 \\ & & & & 0.900875 & -0.040340 \\ & & & & & 4.288783 \end{bmatrix}$$

$$\tilde{\boldsymbol{V}}_{\bar{\boldsymbol{g}}}^{(2)} = \begin{bmatrix} 0.025742 & 0.093216 & -0.000738 & 0.011776 & -0.006746 & 0.005865 \\ & 0.007758 & -0.000300 & 0.003214 & -0.006661 & 0.009726 \\ & & 0.000135 & -0.000363 & 0.000269 & -0.001688 \\ & & & 0.017547 & -0.003343 & 0.014425 \\ & & & & 0.014374 & 0.007653 \\ & & & & & 0.125534 \end{bmatrix}$$

$$\tilde{\boldsymbol{V}}_{\bar{\boldsymbol{g}}}^{(3)} = \begin{bmatrix} -0.004653 & -0.004250 & 0.000487 & -0.004351 & 0.003133 & -0.016605 \\ & -0.003882 & 0.000446 & -0.003974 & 0.002862 & -0.015166 \\ & & -0.000051 & 0.000455 & -0.000328 & 0.001737 \\ & & & -0.004069 & 0.002930 & -0.015523 \\ & & & & -0.002110 & 0.011180 \\ & & & & & -0.059256 \end{bmatrix}$$

By comparison, the submatrix $\boldsymbol{V}_{\bar{\boldsymbol{g}}}$ of the matrix $\boldsymbol{V}_{\boldsymbol{f}}$ is as follows:

$$\boldsymbol{V}_{\bar{\boldsymbol{g}}} = \begin{bmatrix} 0.749256 & 0.006666 & 0.007869 & -0.080288 & -0.130553 & 0.205403 \\ & 1.936482 & -0.018418 & 0.513867 & -0.031020 & -0.369599 \\ & & 0.010773 & -0.023863 & -0.002839 & 0.028431 \\ & & & 1.318745 & 0.139828 & -0.515907 \\ & & & & 0.918900 & -0.043385 \\ & & & & & 4.374648 \end{bmatrix}$$

For this example, the entries of the matrix $\boldsymbol{V}_{\bar{\boldsymbol{g}}}$ closely resemble the entries of $\tilde{\boldsymbol{V}}_{\bar{\boldsymbol{g}}}^{(1)}$,

indicating that the first term $\tilde{\boldsymbol{V}}_{\bar{\boldsymbol{g}}}^{(1)}$ tends to dominate the overall sum $\tilde{\boldsymbol{V}}_{\bar{\boldsymbol{g}}}$. The similarity of the two submatrices is also apparent in Section A.3.2.

The numerical components for the covariance matrix $\tilde{\boldsymbol{V}}'_{\hat{\beta},\bar{\boldsymbol{g}}}$ are as follows:

$$\tilde{\boldsymbol{V}}_{\hat{\beta},\bar{\boldsymbol{g}}}^{(1)} = \begin{bmatrix} 0.032499 & -0.049918 & -0.001591 & -0.042805 & -0.014231 & 0.050800 \end{bmatrix}$$

$$\tilde{\boldsymbol{V}}_{\hat{\beta},\bar{\boldsymbol{g}}}^{(2)} = \begin{bmatrix} 0.000276 & 0.000157 & -0.000046 & -0.000228 & -0.000792 & -0.001831 \end{bmatrix}$$

$$\tilde{\boldsymbol{V}}_{\hat{\beta},\bar{\boldsymbol{g}}}^{(3)} = \begin{bmatrix} 0.000037 & 0.000033 & -0.000004 & 0.000034 & -0.000025 & 0.000130 \end{bmatrix}$$

By comparison, the submatrix $\boldsymbol{V}'_{\hat{\beta},\bar{\boldsymbol{g}}}$ is as follows:

$$\boldsymbol{V}'_{\hat{\beta},\bar{\boldsymbol{g}}} = \begin{bmatrix} 0.032845 & -0.050516 & -0.001601 & -0.043207 & -0.014319 & 0.051279 \end{bmatrix}$$

For this example, the entries of the matrix $\boldsymbol{V}'_{\hat{\beta},\bar{\boldsymbol{g}}}$ closely resemble those of $\tilde{\boldsymbol{V}}_{\hat{\beta},\bar{\boldsymbol{g}}}^{(1)}$, indicating that the term $\tilde{\boldsymbol{V}}_{\hat{\beta},\bar{\boldsymbol{g}}}^{(1)}$ dominates the sum $\tilde{\boldsymbol{V}}'_{\hat{\beta},\bar{\boldsymbol{g}}}$. It should be noted that for this example, there are only six strata and each stratum has fairly large sample size (median size 123) with a reasonable number of events (median number 47) occurring in each stratum.

Dichotomous Outcome: Osteoarthritis Data

For the osteoarthritis data example, there are ten strata with stratum sizes ranging from 10 to 45, with the median stratum size being 20. The components of the covariance matrix $\tilde{\boldsymbol{V}}_{\bar{\boldsymbol{g}}}$ for the WOMAC outcome is as follows:

$$\tilde{\boldsymbol{V}}_{\bar{\boldsymbol{g}}}^{(1)} = \begin{bmatrix} 1.692709 & 0.006789 & -0.123731 & -0.326015 \\ & 0.003603 & -0.027473 & -0.031371 \\ & & 7.480140 & 6.810067 \\ & & & 8.224089 \end{bmatrix}$$

$$\tilde{\boldsymbol{V}}_{\bar{\boldsymbol{g}}}^{(2)} = \begin{bmatrix} 0.467938 & 0.006264 & 0.165081 & 0.016008 \\ & 0.000302 & -0.000680 & -0.001259 \\ & & 1.053048 & 0.563206 \\ & & & 0.507290 \end{bmatrix}$$

$$\tilde{\boldsymbol{V}}_{\bar{\boldsymbol{g}}}^{(3)} = \begin{bmatrix} -0.000966 & -0.000096 & 0.012761 & 0.012455 \\ & -0.000010 & 0.001271 & 0.001241 \\ & & -0.168609 & -0.164569 \\ & & & -0.160625 \end{bmatrix}$$

By comparison, the submatrix $\boldsymbol{V}_{\bar{\boldsymbol{g}}}$ of the matrix $\boldsymbol{V}_{\boldsymbol{f}}$ is as follows:

$$\boldsymbol{V}_{\bar{\boldsymbol{g}}} = \begin{bmatrix} 1.850647 & 0.007463 & -0.135803 & -0.358609 \\ & 0.003937 & -0.029361 & -0.033521 \\ & & 8.194083 & 7.445572 \\ & & & 9.009170 \end{bmatrix}$$

There is some similarity in the covariance matrices $\boldsymbol{V}_{\bar{\boldsymbol{g}}}$ and $\tilde{\boldsymbol{V}}_{\bar{\boldsymbol{g}}}^{(1)}$, but there is also a substantial contribution of some of the entries of $\tilde{\boldsymbol{V}}_{\bar{\boldsymbol{g}}}^{(2)}$ and $\tilde{\boldsymbol{V}}_{\bar{\boldsymbol{g}}}^{(3)}$ to the overall sum $\tilde{\boldsymbol{V}}_{\bar{\boldsymbol{g}}}$.

The numerical components for the covariance matrix $\tilde{\boldsymbol{V}}'_{\hat{\beta},\bar{\boldsymbol{g}}}$ for the WOMAC outcome are as follows:

$$\tilde{\boldsymbol{V}}_{\hat{\beta},\bar{\boldsymbol{g}}}^{(1)} = \begin{bmatrix} -0.016282 & 0.002560 & -0.396712 & -0.448009 \end{bmatrix}$$

$$\tilde{\boldsymbol{V}}_{\hat{\beta},\bar{\boldsymbol{g}}}^{(2)} = \begin{bmatrix} -0.008304 & -0.000415 & 0.001101 & 0.001832 \end{bmatrix}$$

$$\tilde{\boldsymbol{V}}_{\hat{\beta},\bar{\boldsymbol{g}}}^{(3)} = \begin{bmatrix} 0.001090 & 0.000109 & -0.014005 & -0.014055 \end{bmatrix}$$

By comparison, the submatrix $\boldsymbol{V}'_{\hat{\beta},\bar{\boldsymbol{g}}}$ is as follows:

$$\boldsymbol{V}'_{\hat{\beta},\bar{\boldsymbol{g}}} = \begin{bmatrix} -0.016620 & 0.002636 & -0.414590 & -0.468217 \end{bmatrix}$$

Here, it appears as though the matrices $\boldsymbol{V'}_{\hat{\beta},\bar{\boldsymbol{g}}}$ and $\tilde{\boldsymbol{V}}^{(1)}_{\hat{\beta},\bar{\boldsymbol{g}}}$ are similar, indicating that the sum $\tilde{\boldsymbol{V}}'_{\hat{\beta},\bar{\boldsymbol{g}}}$ tends to be dominated by the first component $\tilde{\boldsymbol{V}}^{(1)}_{\hat{\beta},\bar{\boldsymbol{g}}}$.

A.2 Specification of Covariance Matrix $\bar{\boldsymbol{V}}_{\boldsymbol{f}}$ for Compound Vector in Stratified Analysis through Transformation

Let $\boldsymbol{V}_{\bar{\boldsymbol{g}}}$ and $\tilde{\boldsymbol{V}}_{\bar{\boldsymbol{g}}}$ be the lower right $P \times P$ covariance matrices for $\bar{\boldsymbol{g}} = (\bar{\boldsymbol{x}}_{\mathbf{1}} - \bar{\boldsymbol{x}}_{\mathbf{2}})$ per the specifications in Chapter 3 and Appendix Section A.1.1, respectively. Also let $v_{\hat{\beta}}$ be the estimated variance of $\hat{\beta}$ as the cross-products of the DFBETA residuals, and let $\hat{v}_{\beta}$ be the model-based variance estimate of $\hat{\beta}$ from the treatment-only stratified analysis. The matrices $\boldsymbol{V}_{\bar{\boldsymbol{g}}}$ and $\tilde{\boldsymbol{V}}_{\bar{\boldsymbol{g}}}$ can be written as $\boldsymbol{V}_{\bar{\boldsymbol{g}}} = \boldsymbol{T}\boldsymbol{T'}$ and $\tilde{\boldsymbol{V}}_{\bar{\boldsymbol{g}}} = \tilde{\boldsymbol{T}}\tilde{\boldsymbol{T}}'$ where $\boldsymbol{T}$ and $\tilde{\boldsymbol{T}}$ are triangular matrices.

Let $\boldsymbol{D}$ be defined as in (4) where $\mathbf{0}$ is a $P \times 1$ matrix of zeroes.

$$\boldsymbol{D} = \begin{bmatrix} \sqrt{\hat{v}_{\beta}/v_{\hat{\beta}}} & \mathbf{0'} \\ \mathbf{0} & \boldsymbol{T}\tilde{\boldsymbol{T}}^{-1} \end{bmatrix} \tag{4}$$

Then the matrix $\bar{\boldsymbol{V}}_{\boldsymbol{f}} = \boldsymbol{D}\tilde{\boldsymbol{V}}_{\boldsymbol{f}}\boldsymbol{D'}$ provides another specification of a covariance matrix for $\boldsymbol{f}$ which transforms the matrix $\tilde{\boldsymbol{V}}_{\boldsymbol{f}}$ so that $\bar{\boldsymbol{V}}_{\boldsymbol{f}}$ has the quantity $\hat{v}_{\beta}$ as the [1,1] entry and $\boldsymbol{V}_{\bar{\boldsymbol{g}}}$ as the lower right $P \times P$ submatrix.

A.3 Numerical Examples for Different Covariance Specifications

The analyses conducted in Chapter 3 are revisited using the other specifications of the covariance matrix for $\boldsymbol{f}$ presented in Appendix A.1.1 and Appendix A.2.

A.3.1 Dichotomous Outcome: Neurologic Disorder Data

Three specifications of the covariance matrix for $\boldsymbol{f}$ were used to produce the dichotomous outcome analysis results in Table A.3.1 for the 722 patients previously described in Chapter 3. The first estimate, $\boldsymbol{V_f}$, is in (5) and was created through the methodology in Chapter 3.

$$\boldsymbol{V_f} = \begin{bmatrix} 0.029757 & 0.039911 & -0.074802 & -0.002201 & -0.058199 & -0.025258 & 0.069838 \\ & 0.749256 & 0.006666 & 0.007869 & -0.080288 & -0.130553 & 0.205403 \\ & & 1.936482 & -0.018418 & 0.513867 & -0.031020 & -0.369599 \\ & & & 0.010773 & -0.023863 & -0.002839 & 0.028431 \\ & & & & 1.318745 & 0.139828 & -0.515907 \\ & & & & & 0.918900 & -0.043385 \\ & & & & & & 4.374648 \end{bmatrix} \quad (5)$$

The second estimate, $\tilde{\boldsymbol{V}}_{\boldsymbol{f}}$, is in (6) and was created through the methodology in Appendix A.1.1.

$$\tilde{\boldsymbol{V}}_{\boldsymbol{f}} = \begin{bmatrix} 0.029757 & 0.040300 & -0.073789 & -0.002298 & -0.058482 & -0.025186 & 0.068513 \\ & 0.756763 & 0.012887 & 0.007485 & -0.070668 & -0.131936 & 0.190815 \\ & & 1.902509 & -0.017952 & 0.503090 & -0.032487 & -0.365015 \\ & & & 0.010680 & -0.023503 & -0.002825 & 0.027884 \\ & & & & 1.310239 & 0.135858 & -0.504981 \\ & & & & & 0.913140 & -0.021506 \\ & & & & & & 4.355061 \end{bmatrix} \quad (6)$$

The third estimate, $\bar{\boldsymbol{V}}_{\boldsymbol{f}}$, is in (7) and was created through the methodology in

Appendix A.2.

$$\bar{\boldsymbol{V}}_{\boldsymbol{f}} = \begin{bmatrix} 0.029030 & 0.039607 & -0.073863 & -0.002251 & -0.058562 & -0.025006 & 0.068971 \\ & 0.749256 & 0.006666 & 0.007869 & -0.080288 & -0.021773 & 0.205403 \\ & & 1.936482 & -0.018418 & 0.513867 & 0.205904 & -0.369599 \\ & & & 0.010773 & -0.023863 & -0.009212 & 0.028431 \\ & & & & 1.318745 & 0.097392 & -0.515907 \\ & & & & & 0.918900 & -0.043385 \\ & & & & & & 4.374648 \end{bmatrix} \quad (7)$$

All three of the covariance matrix specifications for NPANCOVA produced similar odds ratio estimates, standard errors (with $\bar{\boldsymbol{V}}_{\boldsymbol{f}}$ producing the smallest standard error), and all had a p-value of 0.0002. It should be noted that the standard error using $\bar{\boldsymbol{V}}_{\boldsymbol{f}}$, while not substantially different from the standard error when using $\boldsymbol{V}_{\boldsymbol{f}}$ or $\tilde{\boldsymbol{V}}_{\boldsymbol{f}}$, did show a small decrease relative to these other covariance matrix specifications.

A.3.2 Time-to-Event Outcome: Neurologic Disorder Data

Three specifications of the covariance matrix for $\boldsymbol{f}$ were used to produce the time-to-event analysis results in Table A.3.2 for the 722 patients previously described in Chapter 3. The first estimate, $\boldsymbol{V}_{\boldsymbol{f}}$, is in (8) and was created through the methodology in Chapter 3.

$$\boldsymbol{V}_{\boldsymbol{f}} = \begin{bmatrix} 0.013257 & 0.032845 & -0.050516 & -0.001601 & -0.043207 & -0.014319 & 0.051279 \\ & 0.749256 & 0.006666 & 0.007869 & -0.080288 & -0.130553 & 0.205403 \\ & & 1.936482 & -0.018418 & 0.513867 & -0.031020 & -0.369599 \\ & & & 0.010773 & -0.023863 & -0.002839 & 0.028431 \\ & & & & 1.318745 & 0.139828 & -0.515907 \\ & & & & & 0.918900 & -0.043385 \\ & & & & & & 4.374648 \end{bmatrix} \quad (8)$$

The second estimate, $\tilde{\boldsymbol{V}}_{\boldsymbol{f}}$, is in (9) and was created through the methodology in Appendix A.1.1.

$$\tilde{\boldsymbol{V}}_{\boldsymbol{f}} = \begin{bmatrix} 0.013257 & 0.032776 & -0.049761 & -0.001636 & -0.043033 & -0.015023 & 0.048968 \\ & 0.756763 & 0.012887 & 0.007485 & -0.070668 & -0.131936 & 0.190815 \\ & & 1.902509 & -0.017952 & 0.503090 & -0.032487 & -0.365015 \\ & & & 0.010680 & -0.023503 & -0.002825 & 0.027884 \\ & & & & 1.310239 & 0.135858 & -0.504981 \\ & & & & & 0.913140 & -0.021506 \\ & & & & & & 4.355061 \end{bmatrix} \quad (9)$$

The third estimate, $\bar{\boldsymbol{V}}_{\boldsymbol{f}}$, is in (10) and was created through the methodology in Appendix A.2.

$$\bar{\boldsymbol{V}}_{\boldsymbol{f}} = \begin{bmatrix} 0.013327 & 0.032698 & -0.050611 & -0.001625 & -0.043772 & -0.015142 & 0.050047 \\ & 0.749256 & 0.006666 & 0.007869 & -0.080288 & -0.130553 & 0.205403 \\ & & 1.936482 & -0.018418 & 0.513867 & -0.031020 & -0.369599 \\ & & & 0.010773 & -0.023863 & -0.002839 & 0.028431 \\ & & & & 1.318745 & 0.139828 & -0.515907 \\ & & & & & 0.918900 & -0.043385 \\ & & & & & & 4.374648 \end{bmatrix} \quad (10)$$

The three covariance matrix specifications for NPANCOVA produced similar hazard ratio estimates, standard errors, and p-values. However, the decrease in standard error using $\bar{\boldsymbol{V}}_{\boldsymbol{f}}$ (as compared to $\boldsymbol{V}_{\boldsymbol{f}}$ or $\tilde{\boldsymbol{V}}_{\boldsymbol{f}}$) which was observed in the dichotomous example for the neurologic disorder data did not occur here.

A.3.3 Dichotomous Outcome: Osteoarthritis Data

Three specifications of the covariance matrix for $\boldsymbol{f}$ were used to produce the analysis results in Table A.3.3 for the WOMAC outcome for 218 patients previously described in Chapter 3. The first estimate, $\boldsymbol{V}_{\boldsymbol{f}}$, is in (11) and was created through the methodology

in Chapter 3. The estimated covariance matrices relating to the WOMAC score are as follows:

$$\boldsymbol{V_f} = \begin{bmatrix} 0.080079 & -0.016620 & 0.002636 & -0.414590 & -0.468217 \\ & 1.850647 & 0.007463 & -0.135803 & -0.358609 \\ & & 0.003937 & -0.029361 & -0.033521 \\ & & & 8.194083 & 7.445572 \\ & & & & 9.009170 \end{bmatrix} \tag{11}$$

The second estimate, $\tilde{\boldsymbol{V}}_{\boldsymbol{f}}$, is in (12) and was created through the methodology in Appendix A.1.1.

$$\tilde{\boldsymbol{V}}_{\boldsymbol{f}} = \begin{bmatrix} 0.080079 & -0.023496 & 0.002254 & -0.410012 & -0.460232 \\ & 2.159682 & 0.012956 & 0.054111 & -0.297552 \\ & & 0.003895 & -0.026881 & -0.031389 \\ & & & 8.364579 & 7.208704 \\ & & & & 8.570754 \end{bmatrix} \tag{12}$$

The third estimate, $\bar{\boldsymbol{V}}_{\boldsymbol{f}}$, is in (13) and was created through the methodology in Appendix A.2.

$$\bar{\boldsymbol{V}}_{\boldsymbol{f}} = \begin{bmatrix} 0.076291 & -0.021230 & 0.002279 & -0.394182 & -0.459950 \\ & 1.850647 & 0.007463 & -0.135803 & -0.358609 \\ & & 0.003937 & -0.029361 & -0.033521 \\ & & & 8.194083 & 7.445572 \\ & & & & 9.009170 \end{bmatrix} \tag{13}$$

The three covariance matrix specifications produced similar odds ratio estimates, standard errors, and p-values for the NPANCOVA method. However, the third specification, $\bar{\boldsymbol{V}}_{\boldsymbol{f}}$, resulted in an odds ratio slightly further away from the null, and it had a slightly smaller standard error and p-value than when $\boldsymbol{V_f}$ or $\tilde{\boldsymbol{V}}_{\boldsymbol{f}}$ were used.

Table A.3.1: Stratified Analysis for Dichotomized Outcome from Example 3.3.1 for Neurologic Disorder Data

Method	Log OR	S.E.	OR	95% CI		p-value
				Lower	Upper	
Stratified, Robust Variance[a]	-0.4771	0.1725	0.6206	0.4426	0.8702	0.0057
NPANCOVA Adjusted with $\boldsymbol{V_f}$[b]	-0.5492	0.1470	0.5774	0.4329	0.7702	0.0002
NPANCOVA Adjusted with $\tilde{\boldsymbol{V}}_{\boldsymbol{f}}$[b]	-0.5496	0.1465	0.5772	0.4331	0.7692	0.0002
NPANCOVA Adjusted with $\bar{\boldsymbol{V}}_{\boldsymbol{f}}$[b]	-0.5484	0.1447	0.5779	0.4352	0.7674	0.0002
Stratified, Model-Based Variance	-0.4771	0.1704	0.6206	0.4444	0.8667	0.0049[c]
Conditional Logistic Regression Adjusted[b]	-0.7810	0.2091	0.4579	0.3040	0.6899	0.0002[c]
Unconditional Logistic Regression Adjusted[b]	-0.7895	0.2103	0.4541	0.3007	0.6857	0.0001[c]

[a]Robust variance based on Wei *et. al.*

[b]Adjusted for six continuous baseline covariates

[c]P-value from score test

Table A.3.2: Stratified Analysis for Time-to-Event Outcome from Example 3.3.2 for Neurologic Disorder Data

Method	Log HR	S.E.	HR	95% CI		p-value
				Lower	Upper	
Stratified, Robust Variance[a]	-0.2475	0.1151	0.7808	0.6231	0.9785	0.0316
NPANCOVA Adjusted using $\boldsymbol{V_f}$[b]	-0.3021	0.0942	0.7392	0.6146	0.8891	0.0013
NPANCOVA Adjusted using $\tilde{\boldsymbol{V}}_{\boldsymbol{f}}$[b]	-0.3012	0.0942	0.7399	0.6152	0.8899	0.0014
NPANCOVA Adjusted using $\bar{\boldsymbol{V}}_{\boldsymbol{f}}$[b]	-0.3016	0.0944	0.7400	0.6150	0.8903	0.0014
Stratified, Model-Based Variance	-0.2475	0.1154	0.7808	0.6227	0.9790	0.0316[c]
Cox Adjusted[b]	-0.4829	0.1193	0.6170	0.4884	0.7795	<0.0001[c]

[a]Robust variance based on Wei *et. al.*
[b]Adjusted for six continuous baseline covariates
[c]P-value based on score test

Table A.3.3: Results from Example 3.3.3 for Osteoarthritis Data for the WOMAC Dichotomized Outcome

Method	Log OR	S.E.	OR	95% CI		p-value
				Lower	Upper	
Stratified, Robust Variance[a]	0.6945	0.2830	2.0026	1.1501	3.4872	0.0141
NPANCOVA Adjusted using $\boldsymbol{V_f}$[b]	0.5332	0.2335	1.7044	1.0785	2.6936	0.0224
NPANCOVA Adjusted using $\tilde{\boldsymbol{V}}_{\boldsymbol{f}}$[b]	0.5317	0.2331	1.7018	1.0777	2.6874	0.0226
NPANCOVA Adjusted using $\bar{\boldsymbol{V}}_{\boldsymbol{f}}$[b]	0.5411	0.2276	1.7186	1.0997	2.6834	0.0174
Stratified, Model-Based Variance	0.6945	0.2762	2.0026	1.1655	3.4413	0.0114[d]
Conditional Logistic Regression Adjusted[b]	0.7484	0.3436	2.1135	1.0777	4.1448	0.0276[d]
Unconditional Logistic Regression Adjusted[b,c]	0.7940	0.3535	2.2122	1.1064	4.4232	0.0231[d]

[a] Robust variance based on Wei *et. al.*

[b] Adjusted for sex, Visit 2 continuous age, Visit 1 continuous outcome, Visit 2 continuous outcome

[c] Adjustment for strata through 11 indicator variables

[d] P-values from score test

Appendix B: Chapter 4

B.1 Specification of Covariance Matrix for Compound Vector in NPANCOVA for Multiple Treatment Groups

Let $j = 1, \ldots, n_i$ index patients and $i = 1, \ldots, s$ index treatment groups. Methods for multivariate U-statistics ((Davis and Quade, 1968), (Davis and Quade, 1978), (Kawaguchi, Koch and Wang, 2011)) are used to obtain the approximate covariance matrix of $\hat{\boldsymbol{\beta}}$ and the difference in means of the covariates $(\bar{\boldsymbol{x}}_{\boldsymbol{i}} - \bar{\boldsymbol{x}}_{\boldsymbol{s}})$ for the ith treatment group, the approximate covariance matrix of $(\bar{\boldsymbol{x}}_{\boldsymbol{i}} - \bar{\boldsymbol{x}}_{\boldsymbol{s}})$, and the approximate covariance matrix of $(\bar{\boldsymbol{x}}_{\boldsymbol{i}} - \bar{\boldsymbol{x}}_{\boldsymbol{s}})$ and $(\bar{\boldsymbol{x}}_{\boldsymbol{i'}} - \bar{\boldsymbol{x}}_{\boldsymbol{s}})$ where $i \neq i'$. Let $\boldsymbol{x}_j$ be the vector of covariates for the jth patient. Furthermore, n_i is the number of patients randomized to the ith treatment, and $\sum_{i=1}^{s} n_i = n$. Define an indicator r_{ji} where $r_{ji} = 1$ if patient j is in treatment group i and is 0 otherwise. Also define $u_{jii'} = 0.5$ if patient j is in treatment group i, $u_{jii'} = -0.5$ if patient j is in treatment group i', and is 0 otherwise. Then $\boldsymbol{g}_{ii',jj'}$ for a pair of patients j and j' is defined in (14)

$$\boldsymbol{g}_{ii',jj'} = \frac{n(n-1) r_{ji} r_{j'i'} (u_{jii'} - u_{j'ii'})(\boldsymbol{x}_j - \boldsymbol{x}_{j'})}{2 n_i n_{i'}} \tag{14}$$

When summing over pairs of patients for comparing groups i and i', the following is obtained:

$$\boldsymbol{g}_{ii',j*} = \sum_{j' \neq j} \frac{\boldsymbol{g}_{ii',jj'}}{n-1} = \sum_{j' \neq j} \frac{n r_{ji} r_{j'i'} (u_{jii'} - u_{j'ii'})(\boldsymbol{x}_j - \boldsymbol{x}_{j'})}{2n_i n_{i'}}$$
$$= \begin{cases} \frac{n(\boldsymbol{x}_j - \bar{\boldsymbol{x}}_{i'})}{2n_i} & \text{if } j \text{ from group } i \\ \frac{-n(\boldsymbol{x}_j - \bar{\boldsymbol{x}}_i)}{2n_{i'}} & \text{if } j \text{ from group } i' \\ 0 & \text{if } j \text{ is not from group } i \text{ or } i' \end{cases}$$

Summing over all patients j, the following is obtained for the difference in covariate means between the ith and i'th group:

$$\bar{\boldsymbol{g}}_{ii'} = \sum_j \frac{\boldsymbol{g}_{ii',j*}}{n} = \frac{\bar{\boldsymbol{x}}_i - \bar{\boldsymbol{x}}_{i'}}{2} - \frac{\bar{\boldsymbol{x}}_{i'} - \bar{\boldsymbol{x}}_i}{2} = \bar{\boldsymbol{x}}_i - \bar{\boldsymbol{x}}_{i'}$$

To obtain the variance of $\bar{\boldsymbol{x}}_i - \bar{\boldsymbol{x}}_{i'}$, deviations about the mean for the comparison of the ith and the i'th group are first created:

$$\boldsymbol{g}_{ii',j*} - \bar{\boldsymbol{g}}_{ii'} = \begin{cases} \frac{n(\boldsymbol{x}_j - \bar{\boldsymbol{x}}_{i'})}{2n_i} - (\bar{\boldsymbol{x}}_i - \bar{\boldsymbol{x}}_{i'}) & \\ \quad = \frac{n\boldsymbol{x}_j - n\bar{\boldsymbol{x}}_{i'} - 2n_i\bar{\boldsymbol{x}}_i + 2n_i\bar{\boldsymbol{x}}_{i'}}{2n_i} & \text{if } j \text{ from group } i \\ \quad = \frac{n}{2n_i}(\boldsymbol{x}_j - \bar{\boldsymbol{x}}_i) + \frac{n-2n_i}{2n_i}(\bar{\boldsymbol{x}}_i - \bar{\boldsymbol{x}}_{i'}) & \end{cases}$$
$$= \begin{cases} \frac{-n(\boldsymbol{x}_j - \bar{\boldsymbol{x}}_i)}{2n_{i'}} - (\bar{\boldsymbol{x}}_i - \bar{\boldsymbol{x}}_{i'}) & \\ \quad = \frac{-n\boldsymbol{x}_j + n\bar{\boldsymbol{x}}_i - 2n_{i'}\bar{\boldsymbol{x}}_i + 2n_{i'}\bar{\boldsymbol{x}}_{i'}}{2n_{i'}} & \text{if } j \text{ from group } i' \\ \quad = \frac{-n}{2n_{i'}}(\boldsymbol{x}_j - \bar{\boldsymbol{x}}_{i'}) + \frac{n-2n_{i'}}{2n_{i'}}(\bar{\boldsymbol{x}}_i - \bar{\boldsymbol{x}}_{i'}) & \end{cases}$$

so that

$$\boldsymbol{g}_{ii',j*} - \bar{\boldsymbol{g}}_{ii'} = \begin{cases} \frac{n}{2n_i}(\boldsymbol{x}_j - \bar{\boldsymbol{x}}_i) + \frac{n-2n_i}{2n_i}(\bar{\boldsymbol{x}}_i - \bar{\boldsymbol{x}}_{i'}) & \text{if } j \text{ from group } i \\ \frac{-n}{2n_{i'}}(\boldsymbol{x}_j - \bar{\boldsymbol{x}}_{i'}) + \frac{n-2n_{i'}}{2n_{i'}}(\bar{\boldsymbol{x}}_i - \bar{\boldsymbol{x}}_{i'}) & \text{if } j \text{ from group } i' \\ -(\bar{\boldsymbol{x}}_i - \bar{\boldsymbol{x}}_{i'}) & \text{if } j \text{ not from group } i \text{ or } i' \end{cases}$$

With $\tilde{\boldsymbol{g}}_{ii',j} = \frac{2}{\sqrt{n(n-1)}}(\boldsymbol{g}_{ii',j*} - \bar{\boldsymbol{g}}_{ii'})$, an expression for the variance of $\bar{\boldsymbol{g}}_{ii'}$ ((Davis and Quade, 1968),(Davis and Quade, 1978),(Kawaguchi, Koch and Wang, 2011)) is in (15).

$$\begin{aligned} \boldsymbol{V}_{\bar{\boldsymbol{g}}_{ii'}} &= \frac{4}{n(n-1)} \sum_{j=1}^{n} (\boldsymbol{g}_{ii',j*} - \bar{\boldsymbol{g}}_{ii'})(\boldsymbol{g}_{ii',j*} - \bar{\boldsymbol{g}}_{ii'})' \\ &= \sum_{j=1}^{n} \tilde{\boldsymbol{g}}_{ii',j} \tilde{\boldsymbol{g}}'_{ii',j} \end{aligned} \tag{15}$$

A vector $\boldsymbol{f}_j = (\boldsymbol{r}'_j, \tilde{\boldsymbol{g}}'_{1s,j}, \ldots, \tilde{\boldsymbol{g}}'_{(s-1)s,j})'$ can be created which contains the length $(s-1)$ vector of DFBETA residuals for the jth patient, $\boldsymbol{r}_j = (r_{1s,j}, \ldots, r_{(s-1)s,j})'$, and the deviations about the covariate means for the $s-1$ comparisons of treatment groups with the sth treatment group. The mean vector $\boldsymbol{d} = (\hat{\boldsymbol{\beta}}', (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_s)', \ldots, (\bar{\boldsymbol{x}}_{s-1} - \bar{\boldsymbol{x}}_s)')'$ contains the estimated log hazard ratios $\hat{\boldsymbol{\beta}}$ and differences in means of the P covariates for the comparisons of the $s-1$ treatments with the sth treatment. The variance of $\boldsymbol{d}$ can then be estimated by the $((P+1)(s-1)) \times ((P+1)(s-1))$ matrix $\boldsymbol{V}_{\boldsymbol{d}}$ in (16).

$$\begin{aligned} \boldsymbol{V}_{\boldsymbol{d}} &\approx \sum_j \boldsymbol{f}_j \boldsymbol{f}'_j \\ &= \begin{bmatrix} \sum_j \boldsymbol{r}_j \boldsymbol{r}'_j & \sum_j \boldsymbol{r}_{1s,j} \tilde{\boldsymbol{g}}'_{1s,j} & \cdots & \sum_j \boldsymbol{r}_{(s-1)s,j} \tilde{\boldsymbol{g}}'_{(s-1)s,j} \\ & \sum_j \tilde{\boldsymbol{g}}_{1s,j} \tilde{\boldsymbol{g}}'_{1s,j} & \cdots & \sum_j \tilde{\boldsymbol{g}}_{1s,j} \tilde{\boldsymbol{g}}'_{(s-1)s,j} \\ & & \ddots & \vdots \\ & & & \sum_j \tilde{\boldsymbol{g}}_{(s-1)s,j} \tilde{\boldsymbol{g}}'_{(s-1)s,j} \end{bmatrix} \end{aligned} \tag{16}$$

Here, the covariance matrix for $\hat{\boldsymbol{\beta}}$ is estimated by $\sum_j \boldsymbol{r}_j \boldsymbol{r}'_j$, the covariance matrix for the difference in mean covariates for the ith group and the sth group is approximated by $\sum_j \tilde{\boldsymbol{g}}_{is,j} \tilde{\boldsymbol{g}}'_{is,j}$, the covariance of $\hat{\boldsymbol{\beta}}$ and the difference in mean covariates for the ith group and the sth group is approximated by $\sum_j \boldsymbol{r}_{is,j} \tilde{\boldsymbol{g}}_{is,j}$, and the covariance matrix of the difference in mean covariates for the (i,s) comparison and the (i',s) comparison

$(i \neq i')$ is approximated by $\sum_j \tilde{\boldsymbol{g}}_{is,j}\tilde{\boldsymbol{g}}'_{i's,j}$. For large n, these quantities are estimated by

$$\begin{aligned}
\sum_{j=1}^{n} \tilde{\boldsymbol{g}}_{is,j}\tilde{\boldsymbol{g}}'_{is,j} &= \frac{n(n_i-1)}{(n-1)n_i}\sum_{j\in i}\frac{(\boldsymbol{x}_j-\bar{\boldsymbol{x}}_i)(\boldsymbol{x}_j-\bar{\boldsymbol{x}}_i)'}{n_i(n_i-1)} \\
&\quad + \frac{n(n_s-1)}{(n-1)n_s}\sum_{j\in s}\frac{(\boldsymbol{x}_j-\bar{\boldsymbol{x}}_s)(\boldsymbol{x}_j-\bar{\boldsymbol{x}}_s)'}{n_s(n_s-1)} \\
&\quad + \left[\frac{(n-2n_i)^2}{n_i(n-1)n} + \frac{(n-2n_s)^2}{n_s(n-1)n} + \frac{4(n-n_i-n_s)}{n(n-1)}\right](\bar{\boldsymbol{x}}_i-\bar{\boldsymbol{x}}_s)(\bar{\boldsymbol{x}}_i-\bar{\boldsymbol{x}}_s)' \\
&\quad \approx \boldsymbol{V}_{\bar{\boldsymbol{x}}_i} + \boldsymbol{V}_{\bar{\boldsymbol{x}}_s}
\end{aligned}$$

$$\begin{aligned}
\sum_{j=1}^{n} \boldsymbol{r}_{is,j}\tilde{\boldsymbol{g}}'_{is,j} &= \sqrt{\frac{n}{n-1}}\Big(\sum_{j\in i}\frac{\boldsymbol{r}_j(\boldsymbol{x}_j-\bar{\boldsymbol{x}}_i)'}{n_i} - \sum_{j\in s}\frac{\boldsymbol{r}_j(\boldsymbol{x}_j-\bar{\boldsymbol{x}}_s)'}{n_s}\Big) \\
&\quad + \frac{1}{\sqrt{n(n-1)}}\left[(n-2n_i)\bar{\boldsymbol{r}}_i + (n-2n_s)\bar{\boldsymbol{r}}_s - 2(n-n_i-n_s)\bar{\boldsymbol{r}}_{\notin i,\notin s}\right](\bar{\boldsymbol{x}}_i-\bar{\boldsymbol{x}}_s)' \\
&\quad = \sqrt{\frac{n(n_i-1)}{n_i(n-1)}}\sum_{j\in i}\frac{\boldsymbol{r}_j(\boldsymbol{x}_j-\bar{\boldsymbol{x}}_i)'}{\sqrt{n_i(n_i-1)}} - \sqrt{\frac{n(n_s-1)}{n_s(n-1)}}\sum_{j\in s}\frac{\boldsymbol{r}_j(\boldsymbol{x}_j-\bar{\boldsymbol{x}}_s)'}{\sqrt{n_s(n_s-1)}} \\
&\quad + \frac{1}{\sqrt{n(n-1)}}\left[(n-2n_i)\bar{\boldsymbol{r}}_i + (n-2n_s)\bar{\boldsymbol{r}}_s - 2(n-n_i-n_s)\bar{\boldsymbol{r}}_{\notin i,\notin s}\right](\bar{\boldsymbol{x}}_i-\bar{\boldsymbol{x}}_s)' \\
&\quad \approx \boldsymbol{V}_{\hat{\boldsymbol{\beta}},\bar{\boldsymbol{x}}_i} - \boldsymbol{V}_{\hat{\boldsymbol{\beta}},\bar{\boldsymbol{x}}_s}
\end{aligned}$$

$$\begin{aligned}
\sum_{j=1}^{n} \tilde{\boldsymbol{g}}_{is,j}\tilde{\boldsymbol{g}}'_{i's,j} &= \frac{1}{n_s^2}\sum_{j\in s}\frac{n}{n-1}(\boldsymbol{x}_j-\bar{\boldsymbol{x}}_s)(\boldsymbol{x}_j-\bar{\boldsymbol{x}}_s)' \\
&\quad + \left[\frac{(n-2n_s)^2}{n_s n(n-1)} - \frac{2n_s(n-2n_i)}{n_s n(n-1)} - \frac{2n_s(n-2n_{i'})}{n_s n(n-1)}\right](\bar{\boldsymbol{x}}_i-\bar{\boldsymbol{x}}_s)(\bar{\boldsymbol{x}}_{i'}-\bar{\boldsymbol{x}}_s)' \\
&\quad + \left[\frac{4(n-n_i-n_{i'}-n_s)}{n_s n(n-1)}\right](\bar{\boldsymbol{x}}_i-\bar{\boldsymbol{x}}_s)(\bar{\boldsymbol{x}}_{i'}-\bar{\boldsymbol{x}}_s)' \\
&\quad = \frac{(n_s-1)n}{n_s(n-1)}\sum_{j\in s}\frac{(\boldsymbol{x}_j-\bar{\boldsymbol{x}}_s)(\boldsymbol{x}_j-\bar{\boldsymbol{x}}_s)'}{n_s(n_s-1)} \\
&\quad + \left[\frac{n}{n_s(n-1)} - \frac{4}{n-1}\right](\bar{\boldsymbol{x}}_i-\bar{\boldsymbol{x}}_s)(\bar{\boldsymbol{x}}_{i'}-\bar{\boldsymbol{x}}_s)' \\
&\quad \approx \boldsymbol{V}_{\bar{\boldsymbol{x}}_s}
\end{aligned}$$

Note that in the above expressions, the quantity $(\bar{\boldsymbol{x}}_i - \bar{\boldsymbol{x}}_s)$ has expected value zero on the basis of a valid randomization which would produce comparable covariate distributions

between the treatment groups.

B.2 Covariance Matrices for Multivariate Time-to-Event Outcomes

Extension to multivariate time-to-event outcomes is possible through the creation of the appropriate covariance matrix for the estimated log hazard ratios and the matrix of covariances of the estimated log hazard ratios with the differences in means of covariables between the treatment groups.

Marginal Cox proportional hazards models for each of the K events are fit where treatment is managed as indicators (i.e. with s treatment groups, $z_1, \ldots, z_{s-1}$ represent the $s-1$ indicator variables and the sth group is the referent treatment group). Let $\boldsymbol{R}_i^{(t)}$ be the $n_i \times K$ matrix of the DFBETA residuals for K events for the tth treatment group comparison ($t = 1, \ldots, s-1$) for the n_i individuals from the ith treatment group. Then, via Wei *et. al.* (Wei, Lin and Weissfeld, 1989), $\boldsymbol{R}^{(t)\prime}\boldsymbol{R}^{(t)} = \sum_{i=1}^{s} \boldsymbol{R}_i^{(t)\prime}\boldsymbol{R}_i^{(t)}$ is the robust estimate of the $K \times K$ covariance matrix of $(\hat{\beta}_1^{(t)}, \ldots, \hat{\beta}_K^{(t)})'$ for the tth treatment comparison. With $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1^{(1)}, \ldots, \hat{\beta}_K^{(1)}, \ldots, \hat{\beta}_1^{(s-1)}, \ldots, \hat{\beta}_K^{(s-1)})'$ as the stacked $K(s-1)\times 1$ vector of estimated log hazard ratios, the following $K(s-1) \times K(s-1)$ matrix $\boldsymbol{V}_{\hat{\boldsymbol{\beta}}}$ is in (17).

$$\boldsymbol{V}_{\hat{\boldsymbol{\beta}}} = \begin{bmatrix} \boldsymbol{R}^{(1)\prime}\boldsymbol{R}^{(1)} & \boldsymbol{R}^{(1)\prime}\boldsymbol{R}^{(2)} & \ldots & \boldsymbol{R}^{(1)\prime}\boldsymbol{R}^{(s-1)} \\ & \boldsymbol{R}^{(2)\prime}\boldsymbol{R}^{(2)} & \ldots & \boldsymbol{R}^{(2)\prime}\boldsymbol{R}^{(s-1)} \\ & & \ddots & \vdots \\ & & & \boldsymbol{R}^{(s-1)\prime}\boldsymbol{R}^{(s-1)} \end{bmatrix} \tag{17}$$

Let $\boldsymbol{X}_i = (\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{iP})$ be the $n_i \times P$ matrix of P baseline covariates for the n_i individuals from the ith treatment group. With $\bar{\boldsymbol{x}}_i = (\bar{x}_{i1}, \ldots, \bar{x}_{iP})'$ representing the vector of covariate means for group i, the covariance matrix $\boldsymbol{V}_{\bar{\boldsymbol{x}}_i} = (\boldsymbol{X}_i - \mathbf{1}\bar{\boldsymbol{x}}_i')'(\boldsymbol{X}_i - \mathbf{1}\bar{\boldsymbol{x}}_i')/[n_i(n_i-1)] = \boldsymbol{A'}_i\boldsymbol{A}_i$ is the estimated covariance matrix of the covariate means

for group i and $\mathbf{1}$ is a $n_i \times 1$ vector of ones. Here, $\boldsymbol{A}_i = (\boldsymbol{X}_i - \mathbf{1}\bar{\boldsymbol{x}}_{\boldsymbol{i}}')/\sqrt{n_i(n_i-1)}$.

Let $\boldsymbol{R}_i = (\boldsymbol{R}_i^{(1)}, \ldots, \boldsymbol{R}_i^{(s-1)})$ be the $n_i \times K(s-1)$ matrix of DFBETA residuals for the K events within $s-1$ treatment comparisons for the n_i individuals in the ith treatment group. Then the covariance between $\hat{\boldsymbol{\beta}}$ and $(\bar{\boldsymbol{x}}_i - \bar{\boldsymbol{x}}_s)$ is given as $\boldsymbol{V}_{\hat{\boldsymbol{\beta}},\bar{\boldsymbol{x}}_1} - \boldsymbol{V}_{\hat{\boldsymbol{\beta}},\bar{\boldsymbol{x}}_s} = \boldsymbol{R'}_i\boldsymbol{A}_i - \boldsymbol{R'}_s\boldsymbol{A}_s$.

Let $\boldsymbol{d} = (\hat{\boldsymbol{\beta}}', (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_s)', \ldots, (\bar{\boldsymbol{x}}_{s-1} - \bar{\boldsymbol{x}}_s)')'$ be the $((K+P)(s-1)) \times 1$ vector of estimated log hazard ratios and differences in baseline covariate means for the $s-1$ treatment group comparisons with the sth treatment group. Then $\boldsymbol{V_d}$ is the covariance matrix of $\boldsymbol{d}$ and has the structure in (18).

$$\boldsymbol{V_d} = \begin{bmatrix} \boldsymbol{V}_{\hat{\boldsymbol{\beta}}} & \boldsymbol{V}_{\hat{\boldsymbol{\beta}},\bar{\boldsymbol{x}}_1} - \boldsymbol{V}_{\hat{\boldsymbol{\beta}},\bar{\boldsymbol{x}}_s} & \boldsymbol{V}_{\hat{\boldsymbol{\beta}},\bar{\boldsymbol{x}}_2} - \boldsymbol{V}_{\hat{\boldsymbol{\beta}},\bar{\boldsymbol{x}}_s} & \cdots & \boldsymbol{V}_{\hat{\boldsymbol{\beta}},\bar{\boldsymbol{x}}_{s-1}} - \boldsymbol{V}_{\hat{\boldsymbol{\beta}},\bar{\boldsymbol{x}}_s} \\ & \boldsymbol{V}_{\bar{\boldsymbol{x}}_1} + \boldsymbol{V}_{\bar{\boldsymbol{x}}_s} & \boldsymbol{V}_{\bar{\boldsymbol{x}}_s} & \cdots & \boldsymbol{V}_{\bar{\boldsymbol{x}}_s} \\ & & \ddots & \ddots & \vdots \\ & & & \ddots & \boldsymbol{V}_{\bar{\boldsymbol{x}}_s} \\ & & & & \boldsymbol{V}_{\bar{\boldsymbol{x}}_{s-1}} + \boldsymbol{V}_{\bar{\boldsymbol{x}}_s} \end{bmatrix} \tag{18}$$

Weighted least squares is then used to produce covariate-adjusted estimates $\boldsymbol{b}$ of the log hazard ratios by forcing the difference in means for the covariables to zero. This has invocation for $\boldsymbol{d}$ by fitting the model $\boldsymbol{Z} = [\boldsymbol{I}_{K(s-1)} \quad \mathbf{0}_{K(s-1)\times P(s-1)}]'$ to $\boldsymbol{d}$, where $\boldsymbol{I}_{K(s-1)}$ is a $K(s-1) \times K(s-1)$ identity matrix and $\mathbf{0}_{K(s-1)\times P(s-1)}$ is a $K(s-1) \times P(s-1)$ matrix of zeroes. The covariate-adjusted estimator $\boldsymbol{b}$ is given in (19).

$$\boldsymbol{b} = (\boldsymbol{Z'}\boldsymbol{V_d}^{-1}\boldsymbol{Z})^{-1}\boldsymbol{Z'}\boldsymbol{V_d}^{-1}\boldsymbol{d} \tag{19}$$

and a consistent estimator for the covariance matrix of $\boldsymbol{b}$ is $\boldsymbol{V_b}$ in (20).

$$\boldsymbol{V_b} = (\boldsymbol{Z'}\boldsymbol{V_d}^{-1}\boldsymbol{Z})^{-1} \tag{20}$$

These estimators $\boldsymbol{b}$ have an approximate multivariate normal distribution when the sample sizes for each treatment group are sufficiently large for $\boldsymbol{d}$ to have an approximate multivariate normal distribution, with this usually being sufficient sample size for the $\hat{\boldsymbol{\beta}}^{(t)}$ to have approximate multivariate normal distributions through enough events for each $k = 1, \ldots, K$ in a non-redundant manner. As the differences in means of the covariables between treatment groups are expected to be zero due to randomization, the extent of random imbalances between groups for the means of the covariates may be evaluated through Q_0 in (21).

$$Q_0 = (\boldsymbol{d} - \boldsymbol{Z}\boldsymbol{b})' \boldsymbol{V}_{\boldsymbol{d}}^{-1} (\boldsymbol{d} - \boldsymbol{Z}\boldsymbol{b}) \tag{21}$$

This criterion approximately has the chi-squared distribution with $P(s-1)$ degrees of freedom.

Homogeneity of the adjusted log hazard ratios in $\boldsymbol{b}$ across the K event outcomes (but within a treatment group comparison) can have assessment with the criterion $Q_{h,b}$ in (22).

$$Q_{h,b} = \boldsymbol{b}' \boldsymbol{C}' (\boldsymbol{C} \boldsymbol{V}_{\boldsymbol{b}} \boldsymbol{C}')^{-1} \boldsymbol{C} \boldsymbol{b} \tag{22}$$

where $\boldsymbol{C} = \boldsymbol{I}_{s-1} \otimes [\boldsymbol{I}_{(K-1)} \quad -\mathbf{1}_{(K-1)}]$, and $\otimes$ represents the Kronecker product such that every element in $\boldsymbol{I}_{s-1}$ multiplies the matrix $[\boldsymbol{I}_{(K-1)} \quad -\mathbf{1}_{(K-1)}]$. This criterion approximately has the chi-squared distribution with $(K-1)(s-1)$ degrees of freedom.

Bibliography

Breslow, Norm E and Nicholas E Day. 1980. *Statistical Methods in Cancer Research, Volume I: The Analysis of Case-Control Studies*. International Agency for Research on Cancer, Lyon.

Cain, Kevin C. and Nicholas T. Lange. 1984. "Approximate Case Influence for the Proportional Hazards Regression Model with Censored Data." *Biometrics* 40:493–499.

Cox, D. R. 1972. "Regression Models and Life Tables." *Journal of the Royal Statistical Society, Series B* 34:187–220.

Davis, C. E. and Dana Quade. 1968. "On comparing the correlations within two pairs of variables." *Biometrics* 24:987–995.

Davis, C. E. and Dana Quade. 1978. "U-statistics for skewness and symmetry." *Communications in Statistics - Theory and Methods* 7:413–418.

Efron, Bradley. 1977. "The Efficiency of Cox's Likelihood Function for Censored Data." *Journal of the American Statistical Association* 72:557–565.

Fieller, E.C. 1954. "Some problems in interval estimation." *Journal of the Royal Statistical Society, Series B* 2:175–185.

Jiang, Honghua, James Symanowski, Sofia Paul, Yongming Qu, Anthony Zagar and Shengyan Hong. 2008. "The type I error and power of non-parametric logrank and Wilcoxon tests with adjustment for covariates–A simulation study." *Statistics in Medicine* 27:5850–5860.

Kawaguchi, Atsushi, Gary G. Koch and Xiaofei Wang. 2011. "Stratified Multivariate Mann-Whitney Estimators for the Comparison of Two Treatments with Randomization Based Covariance Adjustment." *Statistics in Biopharmaceutical Research* 3:217–231.

Koch, Gary G. and Catherine M. Tangen. 1999. "Nonparametric Analysis of Covariance and its Role in Noninferiority Clinical Trials." *Drug Information Journal* 33:1145–1159.

Koch, Gary G., Catherine M. Tangen, J-W. Jung and Ingrid A. Amara. 1998. "Issues for Covariance Analysis of Dichotomous and Ordered Categorical Data from Randomized Clinical Trials and Non-parametric Strategies for Addressing Them." *Statistics in Medicine* 17:1863–1892.

Koch, Gary G., Ingrid A. Amara, Gordon W. Davis and Dennis B. Gillings. 1982. "A Review of Some Statistical Methods for Covariance Analysis of Categorical Data." *Biometrics* 38:563–595.

Kong, Lan, Gary G. Koch, Thomas Liu and Huei Wang. 2005. "Performance of some multiple testing procedures to compare three doses of a test drug and placebo." *Pharmaceutical Statistics* 4:25–35.

Lacomblez, Lucette, Gilbert Bensimon, P. Nigel Leigh, Phillippe Guillet and Vincent Meininger. 1996. "Dose-ranging study of riluzole in amyotrophic lateral sclerosis." *Lancet* 347:1425–31.

LaVange, Lisa M., Todd A. Durham and Gary G. Koch. 2005. "Randomization-Based Nonparametric Methods for the Analysis of Multicentre Trials." *Statistical Methods in Medical Research* 14:281–301.

Mancl, Lloyd A. and Timothy A. Derouen. 2001. "A Covariance Estimator for GEE with Improved Small-Sample Properties." *Biometrics* 57:126–134.

Mantel, Nathan and William Haenszel. 1959. "Statistical aspects of the analysis of data from retrospective studies of disease." *J. natl Cancer Inst* 22:719–748.

McNemar, Quinn. 1947. "Note on the Sampling Error of the Difference Between Correlated Proportions or Percentages." *Psychometrika* 12:153–157.

Miettinen, Olli. 1970. "Estimation of relative risk from individually matched series." *Biometrics* 26:75–86.

Moodie, Patricia F., Benjamin R. Saville, Gary G. Koch and Catherine M. Tangen. 2011. "Estimating Covariate-Adjusted Log Hazard Ratios for Multiple Time Intervals in Clinical Trials using Nonparametric Randomization-Based ANCOVA." *Statistics in Biopharmaceutical Research* 3:232–241.

Pincus, T., Gary G. Koch and T. *et. al.* Sokka. 2001. "A Randomized, Double-Blind, Crossover Clinical Trial of Diclofenac plus Misoprostol Versus Acetaminophen in Patients with Osteoarthritis of the Hip or Knee." *Arthritis and Rheumatism* 44:1587–1598.

Pregibon, Darryl. 1981. "Logistic Regression Diagnostics." *The Annals of Statistics* 9:705–724.

Preisser, John S. and Bahjat F. Qaqish. 1996. "Deletion Diagnostics for Generalised Estimating Equations." *Biometrika* 83:551–562.

Saville, Benjamin R. and Gary G. Koch. 2013. "Estimating Covariate-Adjusted Log Hazard Ratios in Randomized Clinical Trials using Cox Proportional Hazards Models and Nonparametric Randomization Based Analysis of Covariance." *Journal of Biopharmaceutical Statistics* (in press).

Saville, Benjamin R., Lisa M. Lavange and Gary G. Koch. 2011. "Estimating Covariate-Adjusted Incidence Density Ratios for Multiple Time Intervals in Clinical Trials using Nonparametric Randomization-Based ANCOVA." *Statistics in Biopharmaceutical Research* 3:242–252.

Simpson, Edward H. 1951. "The Interpretation of Interaction in Contingency Tables." *Journal of the Royal Statistical Society, Series B* 13:238–241.

Snedocor, G. W. and W. G. Cochran. 1980. *Statistical Methods, 7th ed.* Iowa State University Press.

Stokes, Maura E, Charles S Davis and Gary G Koch. 2012. *Categorical Data Analysis Using SAS®, 3rd Edition.* SAS Institute Inc., Cary, NC.

Storer, Barry E. and John Crowley. 1985. "A Diagnostic for Cox Regression and General Conditional Likelihoods." *Journal of the American Statistical Association* 80:139–147.

Tangen, Catherine M. and Gary G. Koch. 1999*a*. "Complementary Nonparametric Analysis of Covariance for Logistic Regression in a Randomized Clinical Trial Setting." *Journal of Biopharmaceutical Statistics* 9:45–66.

Tangen, Catherine M. and Gary G. Koch. 1999*b*. "Nonparametric Analysis of Covariance for Hypothesis Testing with Logrank and Wilcoxon Scores and Survival-Rate Estimation in a Randomized Clinical Trial." *Journal of Biopharmaceutical Statistics* 9:307–338.

Tangen, Catherine M. and Gary G Koch. 2001. "Non-parametric analysis of covariance for confirmatory randomized clinical trials to evaluate dose-response relationships." *Statistics in Medicine* 20:2585–2607.

Wei, L. J., D. Y. Lin and L. Weissfeld. 1989. "Regression Analysis of Multivariate Incomplete Failure Time Data by Modeling Marginal Distributions." *Journal of the American Statistical Association* 84:1065–1073.